Structure, Folding and Interactions of Membrane-associated Biomolecules studied by NMR

Dissertation
zur
Erlangung der naturwissenschaftlichen Doktorwürde
(Dr. sc. nat.)
vorgelegt der
Mathematisch-naturwissenschaftlichen Fakultät
der
Universität Zürich
von
Jiří Mareš
aus der
Tschechischen Republik

Promotionskomitee
Prof. Dr. Oliver Zerbe
Prof. Dr. John Robinson

Zürich, 2009

Tatble of Contents

Summary

In the course of my doctoral studies I focused on the characterization of the structure and folding of membrane-interacting molecules such as the neuropeptide peptide YY (PYY), glycolipids derived from gp120, or polymyxins and their interactions with the membrane-surface components such as lipopolysaccharides (LPS). Studies of structure and folding were caried out on PYY, a small peptide for which membrane association was postulated by us previously, whereas the study on the glycolipid-cyanovirin and polymyxin-lipopolysaccharid systems focused on the interaction between peptides/proteins and carbohydrates.

Some of the peptides from the neuropeptide Y family are know to adopt a relatively stable and well-defined helical hairpin structure in water. The characteristic hairpin formed by a polyproline helix associated with an α-helix results in a particular type of teriary structure known also as the PP-fold. The latter was first determined by X-ray crystallographic analysis of the avian pancreatic peptide (aPP). In this work we have investigated the significance of contacts between resuides from the α-helix and the polyproline helix for forming tertiary structure. In order to do so we redetermined the structure of PYY in water using an extended set of NOEs complemented by RDCs. These data helped to significantly increase the resolution of the structure, and to propose details important for the particular nature of tertiary interactions. The most remarkable improvement was obtained for the turn region linking these two helices. The tertiary interactions are formed by specific Tyr-Pro contacts, and similar contacts are observed in other peptides of the family, that form the characteristic PP-fold. Furthermore, the importance of these interactions was analysed by substituting them via site-directed mutagenesis. Subsequent characterization of their dynamics was carried out by measuring heteronuclear $^{15}N\{^{1}H\}$-NOE data, as well as chemical shift changes during thermal and solvent induced (un)folding of wild type PYY. The unfolding was followed by tracking the changes of the C-α chemical shifts. The thermal denaturation revealed concerted changes occurring at all atoms observed, and according to the cooperative nature of the changes, classified PYY as a two-state folder. The solvent denaturation was performed by stepwise changing the solvent system from pure water to pure methanol in 10% intervals, while the other conditions were kept constant. In previous work of our group, methanol was shown to mimic well the environment at the water-membrane interface. As an example, it was shown, that the structure of PYY in DPC micelles lacked tertiary contacts and is very similar to the one present in methanol. Therefore, the methanol denaturation resembles structural changes accompanying increasing participation in membrane-surface contacts. This study determined the concentration at which the tertiary structure denatures, but it also revealed rigidification of secondary structures at high methanol concentrations. The complex methanol unfolding curves also revealed a dramatic alteration of proline isomerization dynamics.

The second part of my doctoral work focused on studying a glycolipid, derived from the carbohydrate portion of the HIV glycoprotein gp120, incorporated into DPC micelles as a membrane model. In order to validate recognition of this memrane-incorporated glycolipid model, its interaction with a carbohydrate-recognizing protein was studied by high-resolution NMR spectroscopy. In cooperation with Prof. Schmidt from the University of Konstanz a terminal arm of the high-mannose structure of the HIV surface-envelope protein gp120 was synthesized as the mono- di- and tri-mannose moiety linked to an aliphatic membrane anchor. I have then characterized

the insertion of this glycolipid into DPC micelles using spin-labels and by measuring the changes in translation diffusion rates by NMR techniques.

Cyanovirin (CV-N) is cyanobacterial lectin, which binds the above-mentioned high-mannose arm of gp120, thereby preventing HIV virus cell entry. A ^{15}N isotopically labeled CV-N was used in chemical shift mapping studies to monitor its interaction with the micelle-anchored glycolipid. The results proved that CV-N recognizes properly the micelle-anchored glycolipid. Using DPC-incorporated spin labels, it was further confirmed that upon CV-N binding the glycolipid remains anchored into the DPC micelles. It could be shown, that binding occurs at the same sites as for the free dimannoside. In contrast to hydrophobic tertiary interactions of PYY, these interactions include polar groups, which were not disrupted by the membrane environment. Additionally, upon competition with the better ligand trimannoside, we detected detachment of the DPC-micelle-bound CV-N. We envisage that this system may become useful for the reversible attachment of biomolecules to membrane surfaces.

Lipopolysaccharide (LPS) forms the main constituent of the outer membrane of Gram-negative bacteria. LPS is also known for its toxic effects. Septic shock, with a mortality rate of about 50%, is a result of hyperactivation of the immune system and is caused in 70% of cases by LPS. This negatively charged molecule inevitably interacts with cationic antimicrobial peptides, such as polymyxins (PMXs). The strong interaction leads to disruption of the integrity of the outer membrane and may suppress the septic shock initiated by LPS molecules released from disrupted bacterial membranes. In the last part of this thesis work we investigated the LPS-polymyxin interaction. We have chosen a bacterial strain producing a simple form of LPS. A purification method was developed that consisted of an extraction protocol, used before, and a new HPLC chromatography step in combination with an optimized ternary solvent mixture. This procedure facilitated production of the isotopically labeled and chemically defined compound in high purity. Various techniques of heteronuclear NMR spectroscopy allowed observing its insertion in DPC micelles. Furthermore, characterization of the LPS-polymyxin interaction was performed using commercially available polymyxin-B and -E, as well as polymyxin-M (PMX-M), which was expressed, isolated and purified in house. Chemical shift mapping was performed by measuring carbon-proton HSQC experiments of isotopically labeled LPS and unlabeled peptides. In addition, in a complementary experiment isotopically labeled polymyxin-M and unlabeled LPS were used. The data enabled us to localize the interaction sites between LPS and polymyxins. Additional information was derived from isotope-filtered NOESY experiments using ^{13}C-labeled LPS and unlabeled polymyxins. Since the signals of the majority of atoms involved in the intermolecular interaction cannot be observed due to linebroading caused by exchange effects, NOESY experiments did not provide sufficient information for the determination of the LPS-PMX complex at reasonable detail. We have employed a combination of simulated annealing and molecular dynamics calculation to determine the possible structure of the LPS-PMX complex in presence of micelles. The simulated annealing utilized sparse experimental restrains derived from the isotope-filtered NOESY and generated a large set of conformers. Analysis of the obtained set allowed selection of those structures, where the intermolecular contacts were in agreement with the chemical shift mapping patterns. Further refinement and analysis was performed by molecular dynamics calculation. Both solvent (water) and cosolvent (DPC) molecules were explicitly included in the calculation. In this work, we have prepared and characterized all the constituents of the complex biological interaction, proposed the structure of the complex and characterized the nature of the interacting moieties.

Zusammenfassung

Während meiner Doktorarbeit habe ich mich auf die Charakterisierung der Struktur und der Faltung von membran-interagierenden Molekülen konzentriert, darunter das Neuropeptid YY (PYY), Glycolipide abstammend von gp120, und Polymyxine und ihre Interaktionen mit Membranoberflächenkomponenten wie Lipoposacchariden (LPS). Die Struktur- und Faltungsstudie wurde mit PYY, einem kleinen Peptid, durchgeführt, für welches von uns im Vorfeld Membranassoziation postuliert wurde, während die Untersuchung der Glycolipid-Cyanovirin- und Polymyxin-Liposaccharid-Systeme sich auf die Interaktion zwischen Peptid/Protein und Kohlehydraten konzentrierte.

Einige der Peptide der Neuropeptid Y Familie sind dafür bekannt eine relativ stabile und gut definierte helikale “Haarnadel”-Struktur in Wasser anzunehmen. Die charakteristische “Haarnadel”, die durch eine Polyprolin-Helix assoziiert mit einer α-Helix gebildet wird, resultiert in einem bestimmten Typ von Tertiärstruktur, bekannt als PP-Faltung. Letztere wurde zuerst durch röntgenkristallographische Untersuchung des pankreatischen Peptides aPP in Vögeln bestimmt.

In dieser Arbeit haben wir die Bedeutung von den Kontakten zwischen Resten der α-Helix und der Polyprolin-Helix untersucht, die die Tertiärstruktur bilden. Um dies zu tun, haben wir die bekannte Struktur von PYY in Wasser mittels eines erweiterten Sets von NOEs und RDCs verfeinert. Diese Daten halfen, die Auflösung der Struktur signifikant zu verbessern, und Details vorzuschlagen, die wichtig für die besondere Natur der Tertiärinteraktionen sind. Die bemerkenswerteste Verbesserung wurde für die, die zwei Helices verbindende “turn”-Region erhalten. Die Tertiärinteraktionen werden von spezifischen Tyr-Pro-Kontakten gebildet, und ähnliche Kontakte werden in anderen Peptiden der Familie beobachtet, die die charakteristische ”PP-Faltung” zeigen. Ausserdem wurde die Bedeutung dieser Interaktionen durch Punktmutationen analysiert. Anschliessend wurde die Dynamik durch Messen der heteronuklearen $^{15}N\{^{1}H\}$-NOE Daten, sowie die Änderung während dem thermischem und lösungsmittel-induzierten (Ent)Falten von PYY charakterisiert. Das Entfalten wurde beobachtet anhand der Änderung der chemischen Verschiebung der C-α Atome. Die thermische Denaturierung zeigte konzertierte Änderungen bei allen beobachteten Atomen, und aufgrund der kooperativen Natur der Änderungen, wurde PYY als Zweizustands-Falter klassifiziert. Die Lösungsmitteldenaturierung wurde mittels schrittweisem Ändern des Lösungssystems von reinem Wasser zu reinem Methanol in 10 % -Intervallen durchgeführt, während die anderen Bedingungen konstant gehalten wurden. In früheren Arbeiten unserer Gruppe wurde gezeigt, dass Methanol die Umgebung an der Wasser-Membran-Grenzfläche gut nachahmt. Als ein Beispiel wurde gezeigt, dass die Struktur von PYY in DPC-Micellen keine Tertiärkontakte aufwies und dass sie der in Methanol sehr ähnlich ist. Daher ähnelt die Methanol-Denaturierung strukturellen Änderungen, die zunehmende Beteiligung in Membranoberflächenkontakten begleiten. Diese Untersuchung bestimmte die Konzentration, bei der die Tertiärstruktur denaturiert, aber sie zeigte auch eine rigidere Sekundärstruktur bei hohen Methanolkonzentrationen. Die komplexen Methanol-Entfaltungskurven zeigten auch eine drastische Änderung der Prolin-Isomerisierungsdynamik.

Der zweite Teil meiner Doktorarbeit konzentrierte sich auf die Studie eines Glykolipids, das vom Kohlenhydratanteil des HIV Glykoproteins gp120 abstammt, welches in DPC-Micellen als Membranmodel inkorporiert wurde. Um die Erkennung

dieses membraninkorporierten Glykolipid-Models zu validieren, wurde seine Interaktion mit einem kohlenhydrat-erkennenden Protein mittels hochauflösender NMR-Spektroskopie untersucht. In Kooperation mit Prof. Schmidt von der Universität Konstanz wurde ein terminaler Arm der "high-mannose-structure" des HIV-Oberflächen-Hüllen-Proteins gp120 als Mono-, Di- und Tri-Mannoseteil verbunden mit einem aliphatischen Membrananker synthetisiert. Daraufhin wurde die Einfügung dieses Glykolipids in DPC Micellen mittels "spin labels" und durch die Messung der Änderungen der Translationsdiffusionsraten mit NMR-Techniken charakterisiert.

Cyanovirin (CV-N) ist ein cyanobakterielles Lectin, das den oben erwähnten "high-mannose arm" von gp120 bindet und dadurch den HIV Virus am Eintritt in die Zelle hindert. Ein ^{15}N-isotopenmarkiertes CV-N wurde für "chemical shift mapping"-Studien verwendet, um seine Interaktion mit dem micellenverankerten Glykolipid zu beobachten. Die Ergebnisse bewiesen, dass CV-N das micellenverankerte Glykolipid richtig erkennt. Mit den DPC-inkorporierten "spin labels" wurde ausserdem bestätigt, dass das Glycolipid während der CV-N-Bindung in den DPC Micellen verankert bleibt. Es konnte gezeigt werden, dass die Bindung an den gleichen Stellen wie bei Bindung der freien Dimannose stattfindet. Im Gegensatz zu hydrophoben Tertiärinteraktionen von PYY wurden diese Interaktionen, die polare Gruppen einschliessen, nicht durch die Membranumgebung aufgebrochen. Zusätzlich haben wir unter Kompetition mit dem besseren Liganden Trimannosid die Ablösung des DPC-micellengebundenen CV-N detektiert. Wir stellen uns vor, dass dieses System für die reversible Bindung von Biomolekülen an Membranoberflächen nützlich sein könnte.

Lipopolysaccharide (LPS) bilden den Hauptbestandteil der äusseren Membran von gram-negativen Bakterien. LPS ist auch für seine toxischen Effekte bekannt. Septische Schocks mit einer Sterberate von 50 % sind die Folge einer Hyperaktivierung des Immunsystems, welche in 70 % aller Fälle durch LPS verursacht wird. Dieses negativ geladene Molekül interagiert zwangsläufig mit kationischen antimikrobiellen Peptiden zum Beispiel Polymyxinen (PMXs). Die starke Interaktion führt zur Aufbrechung der Integrität der äusseren Membran und kann den septischen Schock unterdrücken, welcher durch von der zerstörten Bakterienmembran freigewordenen LPS-Molekülen ausgelöst wird.

Im letzten Teil dieser Doktorarbeit haben wir die LPS-Polymyxin-Interaktion untersucht. Wir haben einen Bakterienstamm gewählt, der eine einfache Form von LPS produziert. Es wurde eine Aufreinigungsmethode entwickelt, die aus einem bereits etablierten Extraktionsprotokoll und einem zusätzlichen HPLC-Chromatographieschritt in Kombination mit einer optimierten ternären Lösungsmittelmischung bestand. Diese Prozedur ermöglichte die Produktion der isotopenmarkierten und chemisch definierten Verbindung mit einem hohen Reinheitsgrad. Verschiedene Techniken der heteronuklearen NMR-Spektroskopie erlaubten es, ihre Einfügung in die DPC Micellen zu beobachten. Ausserdem wurde die LPS-Polymyxin-M-Interaktion sowohl anhand von kommerziell erhältlichem Polymyxin-B und –E, als auch Polymyxin-M (PMX-M), welches im Haus exprimiert, isoliert und aufgereinigt wurde, charakterisiert. „Chemical shift mapping" wurde durch Messen von Kohlenstoff-Protonen-HSQC-Experimenten mit isotopenmarkiertem LPS und unmarkierten Peptiden durchgeführt. Zusätzlich, in einem anderen komplementären Experiment wurden isotopenmarkiertes Polymyxin-M und unmarkiertes LPS benutzt. Die Daten erlaubten uns, die Interaktionsstellen zwischen LPS und Polymyxin-M zu lokalisieren. Zusätzliche Information wurde von

isotopengefilterten NOESY-Experimenten mit ^{13}C-markiertem LPS und unmarkiertem Polymyxin erhalten. Da die Signale der meisten Atome, die in die intermolekulare Interaktion involviert sind, wegen durch Austauscheffekte verursachten Linienverbreiterungen nicht beobachtet werden können, konnten die NOESY-Experimente nicht genügend Information für die Bestimmung des LPS-PMX-Komplexes in ausreichendem Detailgrad liefern. Wir haben eine Kombination von simuliertem 'Annealing' und Molekulardynamikberechnungen benutzt, um eine mögliche Struktur des LPS-PMX-Komplexes in Anwesenheit von Micellen zu bestimmen. Das simulierte 'Annealing' benutzte wenige experimentelle Einschränkungen, die von dem isotopengefilterten NOESY abgeleitet wurden und generierte ein grosses Set an Konformeren. Die Analyse des erhaltenen Sets erlaubte eine Selektion der Strukturen, in denen die intermolekularen Kontakte im Einverständnis mit dem 'chemical shift mapping'-Muster waren. Eine weitere Verfeinerung und Analyse wurde durch eine 'molecular dynamics' Simulation durchgeführt. Sowohl die Lösungsmittel- (Wasser) als auch die Kolösungsmittelmoleküle (DPC) wurden explizit in der Simulation eingeschlossen. In dieser Arbeit haben wir alle Bestandteile der untersuchten komplexen biologischen Interaktion hergestellt und charakterisiert, wir haben eine Struktur des Komplexes vorgeschlagen und die Eigenschaften der interagierenden Teile charakterisiert.

Contents

Importance of protein folding and binding events - experimental and computational methods

Although the study of protein folding started several decades ago, the rapid development of the experimental and computational techniques has enabled modern protein folding studies that are capable of delivering information on shorter timescales and with higher resolution and reliability. The NMR techniques have reached the stage, in which these can be routinely used for structural studies of small biomolecules. Sophisticated applications suitable for the protein folding studies are still in rapid development. Computational techniques are witnessing a similarly rapid progress, which is also inevitably coupled to the available computational power. Whereas molecular dynamics simulations of well-behaved stable proteins can be nowadays performed routinely, only rather recently it became possible to observe repeated peptide/protein folding events for small, rapidly folding systems *in silico*. The understanding of forces driving protein folding has also resulted in success in *ab initio* structure prediction. Rather then reviewing the history of protein folding studies, the aim of this introduction is to show the importance of protein folding and to give an overview of modern, high-resolution techniques with an emphasis on recent developments of NMR techniques suitable for protein folding studies. In the second part of this introduction modern computational techniques are discussed. Some of the methods will be presented along with case studies. An emphasis will be on studies in which a combination of experimental and computational techniques was applied.

1. Importance of protein folding and binding

1.1 Introduction

Following Anfinsens seminal work it is now generally accepted, that the three dimensional structure of protein is determined by its amino acid sequence[1]. The understanding of the mechanisms determining the dynamic structure ensemble corresponding to global minimum of Gibbs energy under given conditions is still far from complete[2]. The subtlety of effects determining folding becomes obvious when realizing some of the established principles. For example, with the exception of some structural proteins, dynamics or structural rearrangements are essential for the proteins functionality in cases like enzymes, receptors and signaling cascades. High stability, corresponding to a large free energy difference of the natively folded state with respect to any other state would not allow such dynamics. Experimentally, it is known, that the energy corresponding to unfolding is on the order of a single hydrogen bond energy (see e.g. ref[3].). The accuracy, with which the hundreds of contributions would have to be calculated to precisely predict the structure, presents a clear difficulty. Additional insight comes from the dependence of the structure and the folding pathway on the environment. This is highlighted in the attempts to find optimal folding/refolding buffers and other conditions, which allows the protein to find the native conformation *in vitro*. The following section will describe the importance of protein folding, and give a particular example of the consequences, when protein folding fails. Before, a very brief introduction to the history of protein folding will be presented, together with an overview about the recent opinions.

The History of protein folding research

Protein folding studies go back to the beginning of the 20th century. Reviews by Christopher Dobson around 2004 [4] state that "the mechanism by which even a simple protein could even in principle fold to a specific structure was until recently shrouded in mystery". Many general principles were already described in the eighties and early nineties. Since then, considerable progress in this area provided a deeper insight into the mechanisms of protein folding. Many principles are not only understood, but also utilized for protein folding prediction. Below, principles and hypotheses of forces driving protein folding as emerged will be briefly described in chronological order.

Maybe the most important principle in protein folding is the thermodynamic hypothesis, which has been, in its mature form, discussed by Anfinsen [5]. "This hypothesis states that the three-dimensional structure of a native protein in its normal physiological milieu (solvent, pH, ionic strength, presence of other components such as metal ions or prosthetic groups, temperature, and other) is the one in which the Gibbs free energy of the whole system is lowest; in that is, that the native conformation is determined by the totality of interatomic interactions and hence by the amino acid sequence, in a given environment." This hypothesis is widely accepted, supported by the thermodynamic reversibility of folding of many small single-domain proteins[6]. However, the driving force for this thermodynamical stability was still the main subject of discussion.

Driving forces of protein folding

Electrostatic interactions

The fact, that coulomb interactions are strong and long-reaching, together with the knowledge, that protein unfolding can be achieved by acids and bases led to the proposal of Linderstrom-Lang, that the forces driving folding are of electrostatic nature[7]. The electrostatics of a protein is treated as charges on a sphere with a low dielectric constant in a medium with a higher dielectric constant. The extremes of pH make the protein highly charged and the nonspecific repulsion affects protein stability in a negative manner. Specific charge interactions, such as ion pairing or salt bridging, stabilize the protein. These were considered to make a dominant contribution to protein stability[8, 9]. Theoretical and experimental evidences provided insight into the contribution of ion pairing to protein stability. One of the contradictions was for example the high energy of transfer of an ion pair from aqueous solution to the protein interior with low dielectrics, which would indicate that only surface ion pairs contribute to protein stability[10, 11]. The hypothesis of favorable surface ion pairing, however, stands in contrast to the knowledge, that ions in water are solvated rather then forming stable pairs. This lead to the nowadays commonly accepted opinion, that electrostatic interactions do not represent the main driving force for protein folding.

Hydrogen bonds

Hydrogen bonds occur when two electronegative atoms share a hydrogen atom. It is therefore electrostatic in nature, its energy typically ranges from 5 to 30 kJ/mol[12] and depends on the electronegativity and the orientation of the bonding atoms. It was suggested, that hydrogen bond formation is actually the dominant force of protein folding [8]. These authors – originally Pauling and coworkers – proposed hydrogen bonds between backbone carbonyl and amide groups and accordingly predicted structures of α-helix and β-sheets discovered later. The presence of hydrogen bonds in secondary structure elements together with the destabilizing effect of solvents (e.g. formic acid) competing with hydrogen bonds consistently implicated their importance. It was recognized already in the fifties by Kauzmann[13] however, that hydrogen bonds do not fulfill criteria for being the dominant driving force. Unfolded peptide chains experience similarly favorable hydrogen bonds to water molecules. The hydrogen bonding only favors the internal organization once the compact folded structure has been formed through other forces. Only 11% and 12 % of carbonyl and amide groups in proteins do not participate in hydrogen bond formation. Of those participating in hydrogen bonds, 43% form H-bonds to water, 11% to side chains and 46% to backbone amide groups[14]. Further statistics of hydrogen bonding between possible secondary structure elements showed, that hydrogen bonding does not specifically favor a particular structure. As a conclusion, hydrogen bonds, like common classical electrostatic interactions, do not provide a driving force strong enough. They rather introduce specificity of interactions and discriminate unfavorable contacts.

Hydrophobic interactions

In 1949 Linderstrom-Lang, who originally proposed the importance of electrostatic interactions in proteins, recognized, that the electrostatic interactions are not the principal driving force and suggested, that protein folding is driven by the aversion of water to nonpolar residues [15, 16]. The same principle drives the formation of detergent micelles in water. In 1959 Kauzmann stated that the formation of contacts between hydrophobic residues enables the formation of hydrogen bonds among water molecules thus releasing a corresponding amount of energy[13]. The importance of the hydrophobic mechanism was supported by the finding, that nonpolar solvents denature proteins[17]. Nonpolar solvents reduce the free energy of the unfolded state by solvating the exposed nonpolar amino acids. Further evidence was provided by an increasing number of the solved proteins structures from which it became obvious, that hydrophobic residues are predominantly sequestered in the protein interior.

One contribution of the hydrophobic effect is enthalpic. In this, the solvent has more preferential dispersion interactions with itself rather than with other solvent [18]. Mixing is driven by the gain of entropy. The second contribution to the hydrophobic effect is a forced ordering of water molecules on the hydrophobic solute surface. Also this effect is entropic. The local ordering of water molecules is temperature dependent, which makes the free energy of hydrophobic interaction non linearly temperature-dependent, since the heat capacity of this system is temperature dependent. It was shown that the enthalpy of mixing oil in water is generally small and sometimes even negative at 25°[6] leading to the conclusion that the major contribution of oil aversion to water is caused by the above mentioned water ordering. The hydrophobic effect is now recognized as the main driving force in protein folding. The fact that the crucial contribution is provided by the entropic term of the free energy, and that the main term is associated with the solvent rather than the protein makes protein folding difficult to understand. The strong force opposing folding is also entropic in nature, since the protein looses considerable degrees of freedom upon folding. This interplay of counteracting entropic terms, which are difficult to calculate, is far less intuitive than e.g. the coulomb interaction of point charges.

Mechanism of protein folding

Besides the identification of the driving force, the mechanism of folding as a sequence of events transforming an extended polypeptide chain to the folded protein has to be described.

The terminology used to describe the key events and principles in protein folding will be described briefly. It has to be stressed, that this discussion is based on systems of small single-domain proteins, which reversibly fold *in vitro*. Moreover, one should note that the folding process is described after a transfer of such a protein from - typically - high concentrations of denaturant to "native" conditions. Therefore the described situation hardly resembles the situation *in vivo*. However, this does not affect the final folded structure, which according to Anfinsen is independent from the folding pathway. The obtained structure presents the lowest free energy state of the protein, whereas the observed folding pathway strictly refers only to the experimental conditions. Important theories on protein folding are depicted and described in figure A below:

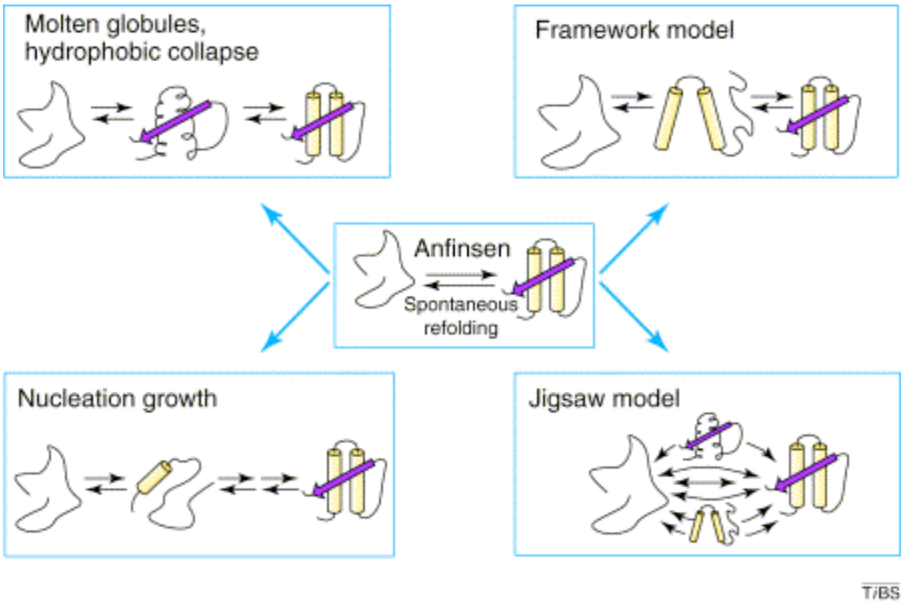


Figure A: Early models for mechanisms of folding. Anfinsen's original experiments demonstrated that proteins fold spontaneously and reversibly into their native conformation. The *nucleation growth*[19] model proposed that residues adjacent in sequence form a nucleus from which the native structure then develops in a sequential manner. By contrast, the *framework model*[20, 21] suggested that local elements of secondary structure form first and that these then dock into the native tertiary structure of the protein, possibly by a *diffusion–collision* mechanism. In the *hydrophobic collapse* model[22], a protein buries its hydrophobic side chains from solvent early during folding, forming a collapsed intermediate or molten globule species, from which the native state develops by searching within this conformationally restricted state. Finally, the *jigsaw model*[23] suggests that each protein molecule could fold by a different path[21]. Although it is now clear that there is not a single sequential folding route, as was implicit in some of these early models, features of these models are relevant today in the context of energy landscape models of folding. Adapted from ref.[25]

A more modern scheme has the form of a funnel-like energy landscape for protein folding, as depicted in figure B. The energy landscape theory has its roots in the statistical mechanics of glasses and phase transitions[24]. The funnel shape signifies the decreasing number of configurations during converging folding pathways. The number of configurations is an approximation, which originates from lattice modeling using discrete states. Other common features, like on-pathway folding intermediates or off-pathway trapped states, can be included into the scheme.

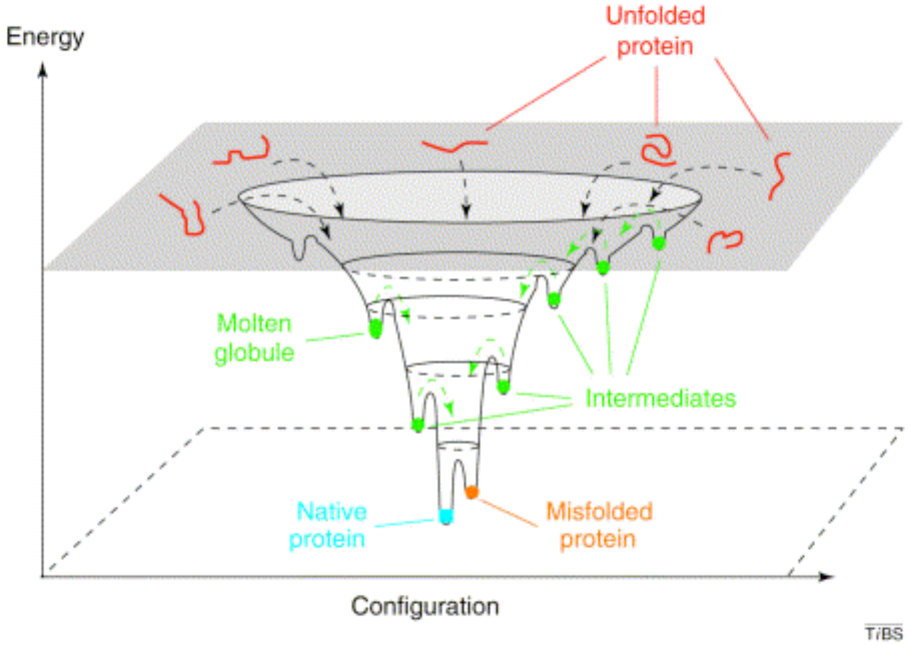


Figure B: Schematic diagram of a folding energy landscape. Denatured molecules at the top of the funnel might fold to the native state by a myriad of different routes, some of which involve transient intermediates (local energy minima) whereas others involve significant kinetic traps (misfolded states). For proteins that fold without populating intermediates, the surface of the funnel would be smooth. Reproduced from ref.[25]

The coordinates, here with rather phenomenological information, often have a more physical content. The number of native-like contacts can serve as a collective variable. An example of such a more concrete scheme is shown and described in figure C

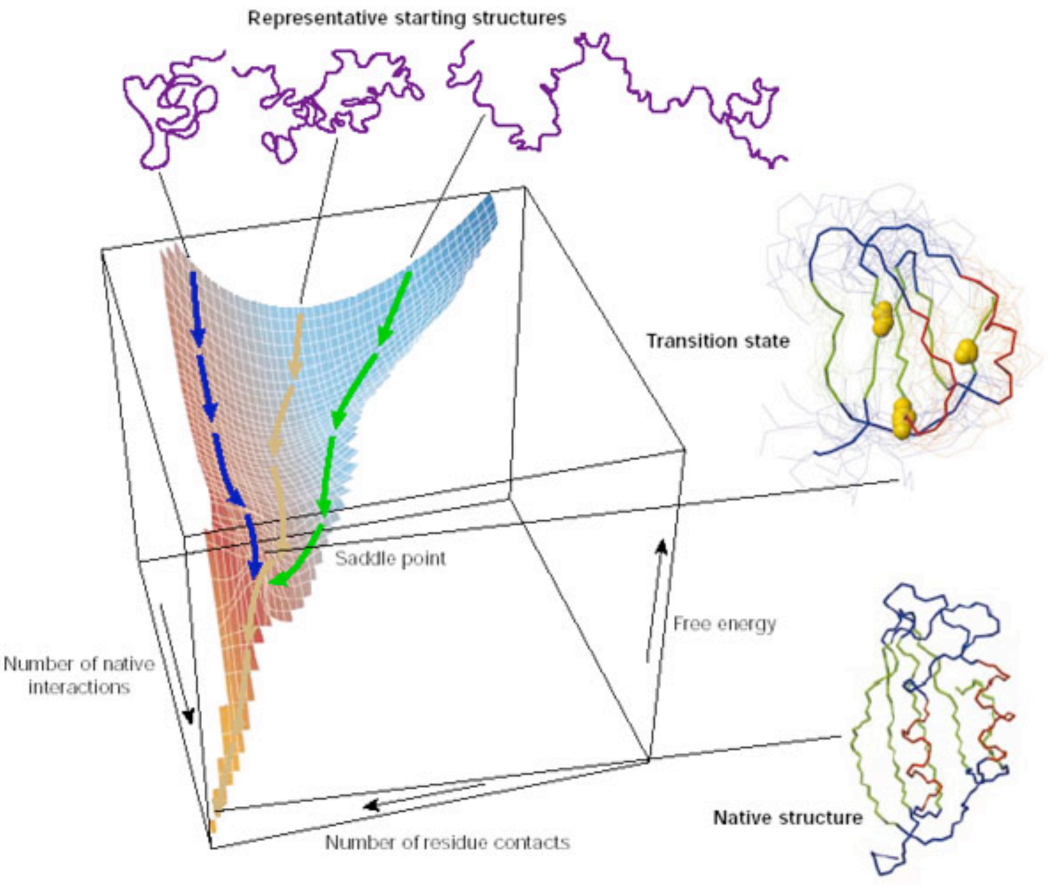


Figure C: The surface is derived from a computer simulation of the folding of a highly simplified model of a small protein. The surface 'funnels' the multitude of denatured conformations to the unique native structure. The critical region on a simple surface such as this one is the saddle point corresponding to the transition state, the barrier that all molecules must cross if they are to fold to the native state. Superimposed on this schematic surface are ensembles of structures corresponding to different stages of the folding process. The transition state ensemble was calculated by using computer simulations constrained by experimental data from mutational studies of acylphosphatase. The yellow spheres in this ensemble represent the three 'key residues' in the structure; when these residues have formed their native-like contacts the overall topology of the native fold is established. The structure of the native state is shown at the bottom of the surface; at the top are indicated schematically some contributors to the distribution of unfolded species that represent the starting point for folding. Also indicated on the surface are highly simplified trajectories for the folding of individual molecules. Reproduced from ref.[26]

1.2 Importance of protein folding

The importance of understanding of protein folding can be seen from many aspects. Proteins are responsible for almost all specific processes in living systems, for example enzymatic reactions, signaling or regulations. A prerequisite to understand these processes is the knowledge of structure and function of proteins. Ability to model protein folding including interactions with its interacting partners can be directly utilized for design of drugs interfering with these processes. Protein folding can be viewed as interactions of the protein with itself allowing simultaneous treatment of protein folding and interactions with other molecules (see e.g. ref.[27]). From this point of view, static pictures of proteins as obtained from X-ray crystallography are rather restricted in information content, and it is therefore not surprising that rational computer drug design still “only helps” in the drug development. Beside the practical aspects, understanding protein folding is a challenging task which will dramatically increase the quality of human knowledge.

1.3 Protein binding and folding

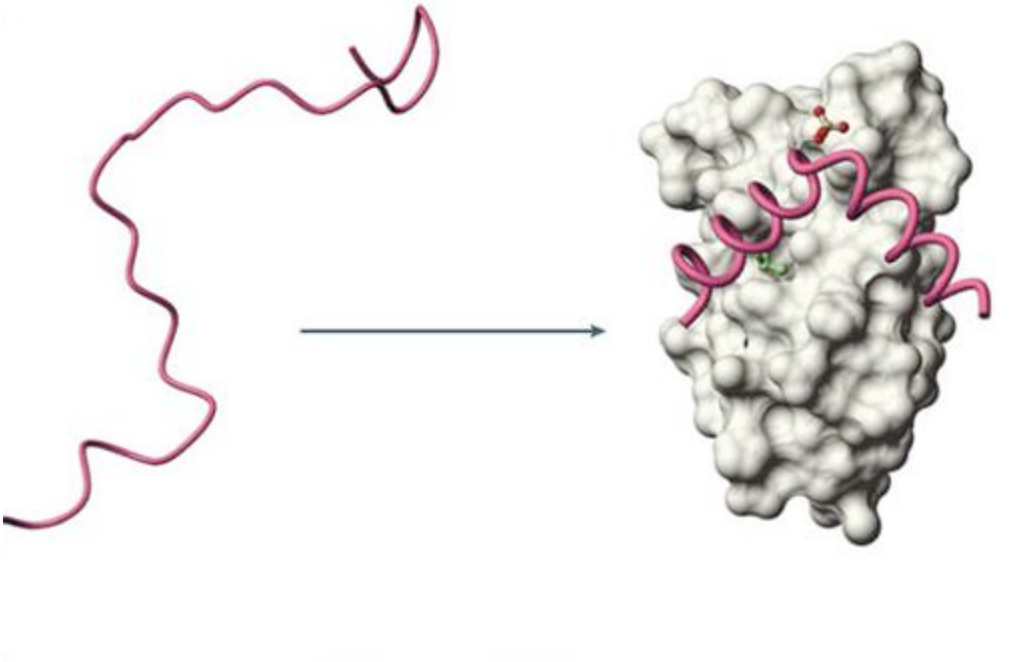


Figure 1: Coupled folding and binding is the process in which an intrinsically disordered protein, or region of a protein, folds into an ordered structure concomitant with binding to its target. On this example, the phosphorylated kinase-inducible domain (pKID) of the transcription factor cyclic-AMP-response-element-binding protein (CREB) is unstructured when it is free in solution, but it folds on forming a complex with the KID-binding (KIX) domain of CREB-binding protein (CBP). Adapted from ref.[28].

The most common hypothesis of protein-ligand interaction is “induced fit”[29], in which the initial interaction between binding partners induces a conformational change in the protein. This can include “fly-casting” effect, where the disordered region binds weakly and non-specifically to its binding target. The other limiting situation of binding is “conformational selection”. In this model, it is assumed, that the free protein samples the variety of low population conformations and the binding partner interacts preferentially with one of these high energy conformers. Recently, this model has been supported by combined NMR relaxation experiments, residual dipolar coupling (RDC) measurement and molecular dynamics (MD) performed on ubiquitin, showing that unbound ubiquitin samples conformations nearly identical to its 46 known bound forms[30].

Protein folding and binding can be understood as analogous processes, in which the protein searches for favorable intramolecular or intermolecular interactions on the energy landscape. This can be probably most prominently seen on the coupled binding and folding mechanism for **intrinsically disordered proteins** (IDPs) or for proteins with unstructured domains. Disordered protein regions show low content of bulky hydrophobic amino acid residues, such as Leu, Ile, Val, Phe, normally forming the stable hydrophobic protein core. Instead, they are rich in particular polar or charged amino acids (Gln, Ser, Pro, Glu, Lys). The functions of these IDPs include regulation processes of transcription and translation, cellular signal transduction, regulation of the self-assembly of large complexes like ribosomes, and chaperone function in folding of other proteins or RNA[27, 28, 31]. Those are involved in signal transduction part of phosphorylation cascades[31]. Their regulatory and signaling function is found with higher occurrence in organisms with higher complexity[28]. For the same reason these unstructured proteins are often associated with cancer and especially with "folding diseases". To this category belong prion neurodegenerative diseases, e.g. Parkinson's disease[32] or "polyglutamine" diseases such as Huntington's disease[33]. The list of human diseases would continue with Creutzfeldt Jacobs Disease (CJD) and Alzheimer disease associated often with the population aging. Bovine Spongiform Encephalopathy (BSE), scrapie affecting sheep or, generally, Transmissible Spongiform Encephalopathies (TSE) are examples of diseases affecting domestic animals[34]. They are directly caused by protein misfolding and have attracted much attention recently.

To appreciate how fragile the folding process is, we should be aware of the chaperon machinery assisting in protein folding[35]. On the opposite end of the protein turnover, much attention is paid to destroy misfolded proteins by ubiquitinilation and subsequent proteolysis[36]. Misfolded proteins are recognized by exposure of hydrophobic residues[28], which in native structures are mostly buried in the hydrophobic core. This is also the key moment for the unwanted interactions to occur, provided the above-mentioned control fails. Hydrophobic parts of misfolded proteins can readily interact with each other and form aggregates, which under physiological conditions are irreversibly insoluble and (therefore) not accessible to degradation processes. It is known, that basically for any protein, conditions under which this fibrillation/polymerization occurs, can be found. This indicates that it is an intrinsic property of proteins. There are factors in protein structure, which increase the propensity of proteins to form fibrils, such as repetitive sequences and low amino acid variability. It is recognized that current proteins are carefully evolutionary selected to be stable in its monomeric (or specifically oligomeric) state[34]. Prion proteins responsible in pathogenic cases, e.g. in BSE, are more prone to misfolding, especially mutants with longer repetitive sequences. The same holds for Alzheimer disease and most likely for other types of these diseases[34]. The progress in understanding of the protein folding/misfolding has led recently to structure-based drug developments against Parkinson disease[37].

2. Experimental methods to study protein folding

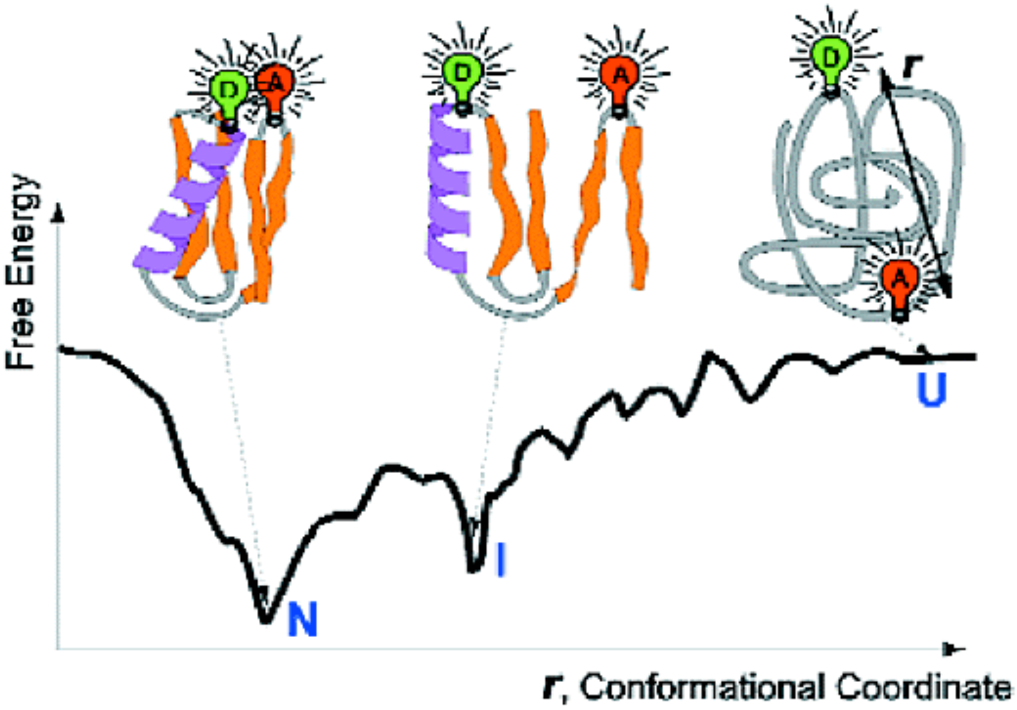


Figure 2: Energy landscape for the folding reaction. A cartoon of free energy as a function of distance *r* between the fluorescence donor and acceptor. The rugged energy landscape has a funnel shape. **U**, the unfolded state; **I**, an intermediate state; **N**, the folded state. Reproduced from ref.[38].

In principle, any method able to capture change upon protein folding can be used to study folding processes. Apart from the NMR related methods, fluorescence methods will be briefly mentioned.

The fluorescence spectrum of Trp is dependent on the polarity of its environment. It can be therefore observed, if the Trp is partitioned into a hydrophobic core, into a membrane environment or exposed to the solution[39].

Fluorescence energy resonance transfer (FRET) is one of the advanced techniques[40]. It is often used to measure the average (or static) distance between fluorescence donor and acceptor. For such a measurement, incorporation of artificial fluorescence probes (donor and acceptor) is needed. Care must be taken that these probes do not disturb the equilibrium of the system. Such a measurement gives simple but valuable information about distance changes during folding. Fluorescence polarization anisotropy (FPA) of a probe rigidly attached to the studied part of molecule is related to overall tumbling, which can differ dramatically upon folding or especially binding, thus giving complementary information.

Recently, single molecule fluorescence techniques have been developed and have become widely used in structural studies. They carry comparably more information than techniques studying the ensemble since the events during folding and unfolding are not synchronized in the ensemble. Through time-resolved, single-molecule (Single Pair FRET, spFRET) or (Single Molecule FPA, smFPA) dynamic effects can be readily monitored[38, 41]. spFRET thus monitors on-line the distances of attached donor and acceptor moieties during folding (see figure 1).

Electron spin resonance is briefly introduced in section 2.2.9. UV circular dichroism spectroscopy is sensitive to formation of secondary structures and can therefore capture their formation. Differential scanning micro-calorimetry is a method suitable for measuring the thermodynamic changes associated with the studied processes, here protein (un)folding and binding. The last two methods do not generally provide atomic resolution data, and will not be introduced in more details.

2.1 NMR related methods

The biggest advantage of NMR spectroscopy is the ability to deliver simultaneously many atomic details of dynamic process associated with folding. The drawback, on the other hand, is the time needed to record the NMR spectrum which, in case of a 2D spectrum, takes usually at least several minutes to record, thus allowing the folding process to be followed in the native conditions only in exceptionally slow cases. An important NMR-related method, for which the time limitation is not determined by the time needed to record the NMR spectrum, is the quench-flow deuterium exchange method. In other cases, it is necessary to find a steady-state condition, in which the average protein population gives information about some stages of folding. This can be done by many methods shifting the equilibrium to states, which are low populated under normal conditions. Namely, elevated temperatures, various denaturant (alcohol) concentrations or high pressure can serve this purpose. Relaxation dispersion can be used to observe dynamics or small rearrangements in the time span of milliseconds. This is especially suited for the study of intrinsically unstructured parts of proteins, rather small rearrangements during enzymatic reaction, and “walks” to excited conformational states of proteins. If suitably arranged by shifting the population along the folding coordinate, it can serve for establishing the trajectory of the folding pathway.

The contribution of NMR to protein dynamics and folding was initiated by studies of Dobson and Carplus using simple methods of 1D NMR spectroscopy, in which for a limited set of resonances, e.g. methyl resonances, resolution was sufficient to follow structural changes with good time resolution[42, 43].

2.1.1 Quench-flow hydrogen/deuterium exchange

The biggest (early) contribution to protein folding as studied by NMR resulted from the Quenched Deuterium Exchange experiment[44, 45]. It provides simultaneously information on the kinetic folding process at residue resolution. The amide protons are labeled at multiple time points during the folding process according to their solvent accessibility at corresponding stage of folding.

In the original version of this experiment, the protein is firstly unfolded by use of denaturant. It is then rapidly diluted into deuterated refolding buffer. The folding is then allowed to proceed for a variable time interval before a 50ms pulse of non-deuterated water is applied, thus labeling amide moieties that during the pulse interval are not shielded from the solvent. Subsequent lowering of pH effectively stops the D/H exchange and the folding is allowed to complete. After refolding is complete a spectrum of the native protein is recorded and the measured quantity is the extent of exchange of given amide protons. One limitation of this technique is that only those amide protons, for which the exchange rate under the final conditions is lower then $10^{-4}s^{-1}$ can be observed in the subsequent 2D experiment. This allows to use 30-50% of backbone amides reliably in these standard quench-flow experiments. The number mostly corresponds to amide protons that are buried in the protein core, the information about exposed ones is lost.

The quench flow experiment has also been conducted differently, making use of aprotic solvent to considerably slow down the exchange process during the NMR spectra recording. This allows gaining information from virtually all of the amide protons[46]. The protein is unfolded preferably in H_2O solution at low pH (e.g. pH=2.2). The folding process is then triggered by transferring the protein sample to D_2O solution of higher pH value (e.g. pH=6). (This pH value remains in a region of slow amide proton exchange regime.) After a given period, during which the initial phase of folding takes place, the second pH jump to even higher pH value (e.g. pH=10) is performed enabling pH-jump labeling (i.e. all unprotected amide protons are exchanged to deuterium). After a variable period of pulse labeling, the pH is lowered again (pH=6) to stop (quench) the amide proton exchange, and the sample is immediately frozen in liquid nitrogen and lyophilized in order to “freeze” the state of exchange. Samples are then dissolved in DMSO with a controlled amount (0.4%) of D_2O. This causes slow but controlled exchange of the remaining protons to deuterium. Therefore, to obtain the protonation corresponding to the time of quenching, several spectra are recorded in defined time intervals and the decay of individual peaks is fitted to an exponential function. From extrapolation to zero time (the time of dissolving in DMSO) the peak intensity at that time is obtained.

These are the only experiments that allow following the kinetics of fast folding events on the amino acid level. On the contrary, a technique like CD spectroscopy can also monitor the kinetics of folding, but only the average quantity over the whole protein is obtained.

For the investigation of protein folding in more detail, it is necessary to find conditions, under which the partly folded state is present (this is true for the observation of the “visible states”; the capabilities of different techniques will be discussed in section 2.2.4). To find such a condition is often not possible, however there are some cases which can serve as models, like apomyoglobin at pH 4[47, 48]. The results from such model systems are likely to be valid more generally. Apomyoglobin at pH 4 is an equilibrium intermediate, so the emerging question is, what kind of information is actually obtained from such an equilibrium intermediate. Does it correspond to a kinetic intermediate, or can it be assumed that these stabilized/equilibrium intermediates correspond to a specific folding event under *native* conditions? This can be expected for two-state models, where we only shift the population between folded and unfolded state, but must not be generally valid. The independent technique, which can observe the kinetic intermediate, is the above-discussed quench-flow experiment. This is commonly used to show that the kinetic and equilibrium intermediates share similar features. If justified the observation of equilibrium folding intermediates can provide experimental information otherwise not available. Intrinsically unfolded proteins on the other hand are always in some sort of a “folding equilibrium” in their native state, so that their study has virtually no timescale restriction[49].

2.2 Information from NMR experiments

2.2.1 Chemical shifts

Chemical shifts are the most readily and rapidly accessible parameters from NMR measurement. They serve as sensitive probes of the local environment. In this regard, the different nuclei present in proteins report slightly differently. Proton chemical shifts are comparably more sensitive to environment in 3D space. Nitrogen chemical shifts have similar properties. The readily measurable nitrogen chemical shifts are backbone amides. However, the sensitivity to many factors, like hydrogen bonding of corresponding protons, makes them most difficult to interpret nowadays. They therefore mainly serve for increasing the dimensionality of the spectra, allowing distinction of individual amino acid signals and tracking their changes. On the contrary, carbon chemical shifts are sensitive to chemical environment in the context of bonded chemical moieties and the local conformation. This makes carbon chemical shifts applicable to derive simple (from a high resolution structural data point of view) structural predictions. Secondary chemical shifts, the differences of measured values from random coil values, are used to identify the presence of secondary structure elements. In this respect, individual carbon atoms (C', $C\alpha$, $C\beta$) possibly together with $H\alpha$ are evaluated for the independent predictions. The stretches of obtained secondary structures from these different atoms can be compared to see if they support the anticipated picture. In the case of partly unfolded proteins, secondary shifts reveals the presence of residual secondary structure[50]. More detailed information is not generally possible because of the dependence of the chemical shift on many factors. Chemical shifts can, however, be used to directly map structural changes during folding if the folding event is a simple two state model. In this case, the chemical shift corresponding to individual states can usually be measured or extrapolated. Then the chemical shift gives information about the population (corresponding to thermodynamics) and possibly also kinetics of the state exchange in a similar manner to other chemical exchange measurements. More information about chemical exchange spectroscopy relevant to protein folding is given in section 2.2.6.

2.2.2 Relaxation data

There are many ways to obtain protein dynamics from NMR experiments. Various relaxation measurements can deliver information about motions on different timescales. The most widely used measurements are probably heteronuclear NOE of amide backbone resonances and ^{15}N T_1 and T_2 measurements[51, 52]. The methodology has been established on well-structured proteins, but the methods are the same for locally or partly unstructured proteins or folding intermediates. These measurements reflect the relatively fast backbone dynamics (ps-ns). This gives quick evidence as to whether the molecule is rigid, in a molten globule or unfolded state. The values present then valuable information regarding folding intermediates. Furthermore, the comparison of the rigidity between different folding intermediates, or dynamic changes upon binding to the binding partner can be made. Binding usually results in rigidification of otherwise more flexible binding sites, however exceptions can be found, where on the contrary binding leads to higher flexibility[53]. Prerequisites for commonly used model-free Lipari-Szabo analysis, in particular the time-separation of internal motions and overall tumbling by one order of magnitude[54], are not fulfilled for partially folded proteins. Therefore these measurements on their own give only rather qualitative information. A way to utilize the information more specifically and in more detail is to apply a specific motional model for which we can fit motional frequencies present in the system. Another approach leading to the dynamic picture is spectral density mapping.

2.2.3 Spectral density mapping

Spectral density mapping allows calculation of the spectral density function at different frequencies based on measurements of a sufficient number of relaxation parameters without postulating any model[55, 56]. In spin systems that are present in proteins, there are several energy/frequency transitions that can be measured. For 2 spin systems like backbone N-H or C-H, the longitudinal and transverse relaxation rates of N and C can be measured besides the cross-relaxation rate. The relaxation of, for example, anti-phase coherence S_xH_z (S=N or C) depends on different combinations of spectral densities in contrast to R1 and R2 rates or the H-NOE. If a sufficient number of relaxation rates is known, it is possible to analytically compute the spectral densities. Measuring the relaxation data at various magnetic fields leads to better determination of the spectral densities. The frequencies of internal motion as derived from relaxation experiments can be also evaluated using the approach of Lipari-Szabo providing order parameters for the individual residues.

The information on nanosecond time-scale motion of side chains can be obtained in a straightforward way when using selective fractional labeling of the methyl moieties[57, 58]. Such a kind of labeling is nowadays widely used to dramatically increase the resolution in methyl region of larger proteins and thus obtain accurate structural information, which would be otherwise inaccessible for larger (slowly reorienting) proteins. Such fractional labeling circumvents the problems associated with carbon-carbon couplings. A method has been developed to measure several different relaxation parameters for the CH_2D methyl groups (including deuteron, carbon and two protons as the NMR active probes), giving valuable information about the local dynamics[59, 60].

All of these techniques give information on fast and therefore small-scale dynamics. For slower motions comprising larger structural changes, other techniques can be employed. For that, special interest has recently been paid to T_2 relaxation dispersion measurements, which are covered in sections bellow, and to measurements of RDCs [61, 62].

2.2.4 Relaxation dispersion - CPMG

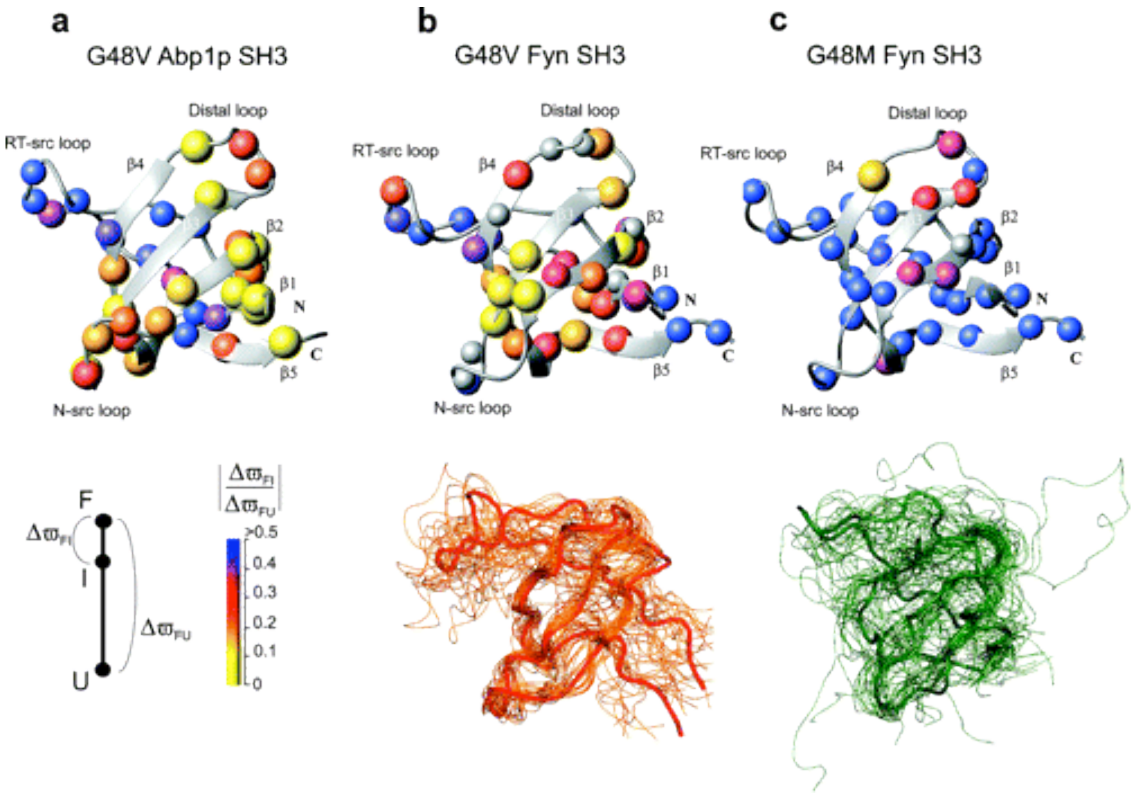


Figure 3: Ratios of ^{15}N chemical shift differences $|\Delta| = \frac{\Delta\omega FI}{\Delta\omega FU}$ for (a) G48V Abp1p SH3,18 (b) G48V Fyn SH3, and (c) G48M Fyn SH3 domains color coded on structures of the wild-type proteins, where each sphere in the diagram is a backbone amide nitrogen position (top plots). Bottom plots in panels b and c show ribbon representations of 25 structures that model the ensembles of the I states of G48V and G48M Fyn SH3 calculated based on experimental Δ ratios. Reproduced from ref.[63]

Relaxation dispersion studies have recently gained much popularity in the field of dynamic studies and, particularly, for probing of so-called invisible excited states[63]. These excited states are formed by a set of conformers with higher Gibbs energy compared to ground state, and according to the Boltzmann distribution are less populated. These experiments enable the study of conformers populated at 0.5% or higher, which are in a rapid exchange with the ground state or other excited states. The fact, that by means of relaxation dispersion one can measure events on a millisecond time-scale makes this technique especially suitable to larger conformational rearrangements. In some cases, it allows following the whole folding process, which fulfills these equilibrium conditions. The extent of information obtained from the study differs from case to case, but in principle kinetic and thermodynamic data are accessible. To obtain a more complete picture of structural changes and kinetic and thermodynamic details, the study can include investigations on the temperature dependence, the effect of applied pressure, and mutations to obtain ϕ-values (see below). A proposed model for folding can be verified from the fit with an experimental data set (two state model, model with a metastable intermediate, etc.). Chemical shifts of these excited states also become accessible, which may directly supply crude structural information. At this point, it should be stressed again, that the chemical shifts corresponding to the particular excited states are not directly visible but will be derived indirectly from the fits.

The idea of relaxation dispersion measurements by Carr-Purcell-Meiboom-Gill (CPMG)[64] is based on the use of the repeated spin echoes (developed in 1950 by Erwin Hahn[65]). The spin echo starts by establishing coherence in the x-y plane. Then, the coherence acquires a phase shift due to chemical shift evolution**.** After a delay τ, an 180° inversion pulse is applied so that after the second τ period the coherence is refocused and the phase shift of the total period is zero. However, if, for example, during the second half of the echo period, the observed spin experiences a different magnetic environment, resulting in a change in the chemical shift, then at the end of the echo period, the overall coherence will be not completely refocused. This frequency change can be due to a walk to one of these excited states. The incompleteness of refocusing is then a function of the chemical shift difference between the ground state and the excited state(s), and the total time period spent in this state. The extent of the decoherence is also the function of the echo length. In practice, echoes are repeated by application of inversion pulses at a frequency ν_{CPMG} (Hz), which then determines the length of individual echoes. Increasing the CPMG frequency results in decrease of the chemical exchange contribution to line-broadening. The resulting R_2 relaxation rate $R_{2,eff}$ is quantified from a fit of the peak intensity vs. the length of the total echo delay time. The relaxation dispersion profile ($R_{2,eff}(\nu_{CPMG})$) allows the extraction of populations of exchanging states, rates of transition between the states and the absolute value of frequency difference (or, equivalently, the chemical shift difference). These quantities are obtained from the fit to a two state exchange or by numerical solving of the Bloch-McConnell equation for a given exchange model[66, 67]. The description so far concerns two isolated spins. In proteins, however, the usually observed unit is formed from (heteronuclear) multiple-spin systems with the most typical one being the backbone amide H-N. In these cases, the situation is more complicated, and requires using different combinations of frequencies. Typically, the exchange model is not known *a priory,* and relaxation dispersion data from only one probe (N or selected C nucleus) are not sufficient to provide discrimination between possible models. Combination of relaxation dispersion profiles from more nuclei allows the choice of the appropriate model that is best suited to describe the observed exchange behavior. Amide groups are the most widely used probes. Cα and carbonyl carbons of the peptide backbone as well as methyl groups of mainly hydrophobic amino acid residues need to be specifically labeled. Recently, it has been shown, that backbone carbonyl carbons, which typically possess sufficiently separated chemical shifts, can be utilized as separated spin system in CPMG experiments for extraction of chemical shifts even in uniformly carbon labeled proteins[68].

Combination of measurements at different magnetic fields, the temperature dependence (and pressure-dependence if possible) can more precisely discriminate between the two-state and more complex folding models. As mentioned above, acquired chemical shifts also help in establishing the complex geometric model of the folding event. Recent developments in the use of chemical shifts in *de novo* structure calculation evoke the possibility for structure calculation of the excited states.

A seminal study has been done on the SH3 domain of **Fyn**[3, 69, 70]. It is a 60 amino acid polypeptide, which has a free energy difference between the ground (folded) state and an unfolded (intermediate) state of only ca. 4-12 kJ/mol corresponding to a 0.5-10% fraction of the excited state. The frequency of unfolding is typically 80 Hz. Moreover, specific single amino acid mutations can further destabilize this protein, which results in an accelerated unfolding rate to 1 kHz and in a decrease of the free energy of unfolding. This makes the unfolded state more easily "visible" from the ground state by the relaxation dispersion measurements. It was also possible to show, that the protein is not a two-state model, as judged from other methods, but that the folding pathway comprises a metastable intermediate. The chemical shift information was utilized in several ways: The difference of chemical shift for a given atom between intermediate state, folded and unfolded states predicts whether the atom in the intermediate structure is in an environment more similar to the folded or the unfolded state. Chemical shifts were also used in biased molecular dynamic protocols utilizing nitrogen chemical shifts. Since nitrogen chemical shifts are mainly sensitive to the 3D environment it can be assumed that the deviation of the nitrogen chemical shift from its ground state value is a measure of unfolding, because it corresponds to the loss of native contacts (see figure 3). The simulated annealing protocol was developed to sample those conformations, for which the measured and calculated chemical shift deviations are in agreement[3]. Since only the nitrogen chemical shift information was used, there is obvious place for improvement.

Another application of relaxation dispersion is its use in a ϕ -value analysis[71]. In this method, a set of single point mutants is prepared to evaluate the influence/role of the residue during the folding. The calculation of the ϕ -value is explained in the figure 4. If the value is close to 1, the mutant is affected in a similar way in both the excited and native states, which means that the residue is involved in native-like contacts also in its excited state. Whereas if the ϕ -value is close to 0, the residue is probably in a disordered region compared to the native folded state. The advantage of the ϕ -value measurement by NMR relaxation dispersion is that accurate kinetic constants between the studied states can be obtained. As a result, the free energy for each mutant can be used to construct the one dimensional energy landscape along the folding process[72, 73].

Thermodynamic information can be obtained from measurement of the temperature dependence of the free energy changes. In this way, both enthalpic and entropic contributions to free energy can be obtained. The role of entropy has been extensively discussed, but no unified view is accepted. This problem is due to the difficulty in computing the entropy contributions. It has been recently stated, that the conformational freedom in unfolded states is by far smaller than usually expected and that the dynamics of a folded protein is larger then commonly expected. Accordingly, the whole entropy difference is smaller. These measurements provide additional details on how the particular folding events are related to entropy changes[53, 74].

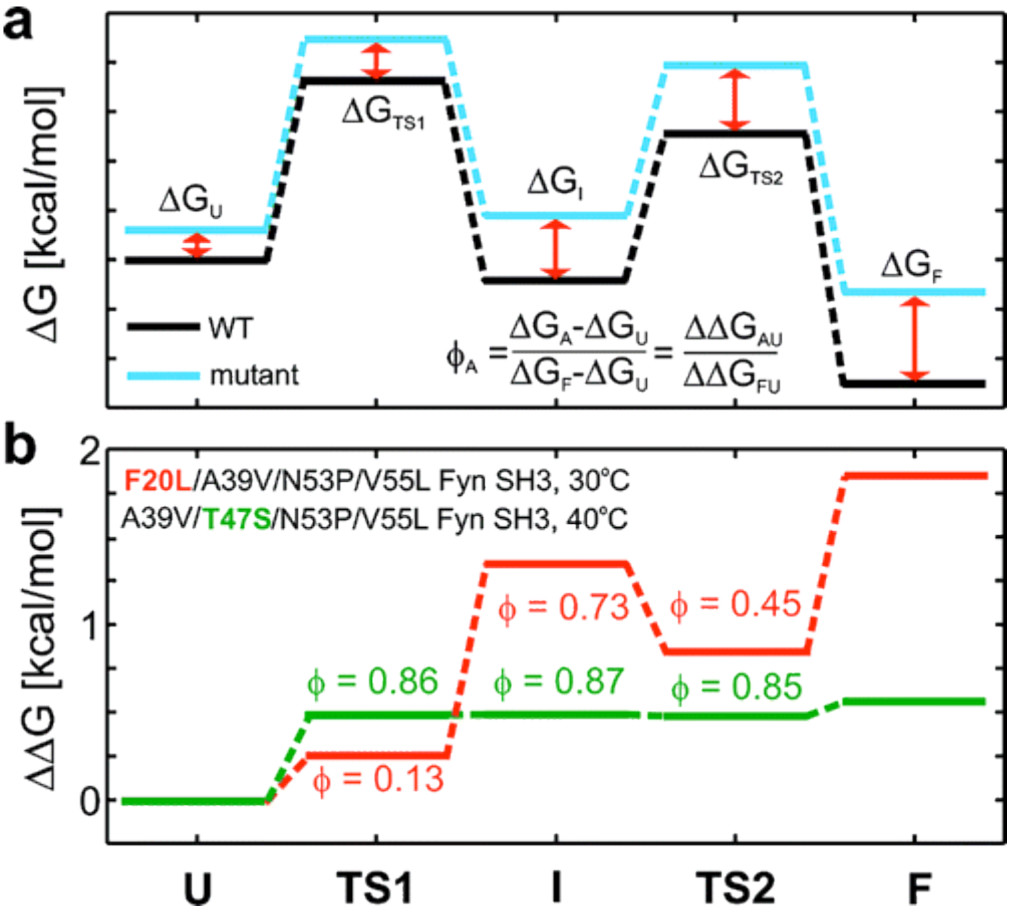


Figure 4: Calculation of φ values for intermediate and transition states along a three-state folding pathway, U ↔ I ↔ F, from changes in free energies upon mutation, ΔG_A (here A denotes one of F, TS1, I, TS2, or U). (b) Changes in free-energy differences $\Delta\Delta G_{AU} = \Delta G_A - \Delta G_U$ for various states along a three-state folding pathway of A39V/N53P/V55L Fyn SH3 caused by T47S (green) and F20L (red) mutations, along with $\varphi_A = \dfrac{\Delta\Delta G_{AU}}{\Delta\Delta G_{FU}}$ values for corresponding intermediate and transition states. Reproduced from ref.[63]

One effect, which has been extensively discussed in the context of protein folding, is the so-called hydrophobic collapse. To monitor the hydrophobic interactions by relaxation dispersion measurements, one can probe the methyl carbon resonances. For this purpose, a special labeling of hydrophobic amino acids (leucine and valine) is needed. The chemical shifts are extracted from the relaxation dispersion data fits. The difference of chemical shift in the transition state with respect to the folded and unfolded state gives qualitative data about the specificity (the state-difference) of the hydrophobic contacts in the transition state, that according the present view follows the hydrophobic collapse. A very interesting phenomenon arises from deuteration of the protein, by which the van-der-Waals (VdW) interactions are weakened. This can have various effects on folding. The obvious general result is that those parts of the structure, for which the VdW contacts are crucial, are destabilized. This may not only affect the folded form but also the transition state thus leading to an increased folding rate[70]. The measured differences of kinetic and thermodynamic properties for folding then serve to quantify the formation of VdW contacts during folding.

The designation “invisible state” is now associated with research done in this field by the lab of Lewis Kay. Arthur Palmer with his group has contributed to the study of protein folding by interpretation of exchange line-broadening or relaxation dispersion CPMG technique several years earlier[59, 75].

2.2.5 R$_{1\rho}$ relaxation dispersion

The relaxation rate of magnetization that is spin-locked in the x-y plane by a continuous radiofrequency irradiation represents an alternative to the CPMG experiment for the measurement of relaxation dispersion caused by chemical exchange. The use of the CPMG technique is limited to rates of chemical exchange (k_{ex}) lower than $10^4 s^{-1}$. This limitation is due to the experimental constraints, namely that a higher frequency of the CPMG pulse train would lead to heating of the sample by the transferred power and could also possible damage the probe. The R$_{1\rho}$ method shifts the limit to $k_{ex} < 10^5 s^{-1}$. It has been proposed to fit simultaneously the CPMG and R$_{1\rho}$ data to increase the range of the accessible effective rf-field and the robustness of the data analysis[76]. The involved methods for the mathematical analysis can be found in ref.[67].

2.2.6 Chemical exchange ZZ-spectroscopy

If the system under study is in the slow exchange regime (milliseconds to seconds), and the unfolded state is significantly populated, meaning that it is visible in the spectrum, it becomes possible to observe the folding process using a zz chemical exchange experiment[69, 77, 78]. The experiment utilizes the small longitudinal relaxation rate of the heteronuclear (N or C) magnetization. Furthermore, for nitrogen, possible artifacts stemming from (N-N NOE) instead of chemical exchange are negligible. For uniformly carbon labeled samples, the undesired C-C NOE can be observed[79]. The exchange constant k_{ex} must not be much smaller than the longitudinal relaxation rate constants on both exchanging sites ($R_{1,a}$, $R_{1,b}$) - typically 0.1-10s. If, for example, folding corresponds to a two-state model, it is possible to observe the N longitudinal magnetization transfer due to the conformational (chemical) exchange. The zero-mixing-time intensities of the diagonal peaks in the exchange spectrum are proportional to populations p_A and p_B. This is practically obtained from the simple HSQC spectrum. The corresponding free energy difference between states A and B is readily available. The crosspeak and diagonal peak intensity depend on the mixing time, which is varied in a series of experiments. The analysis of the data using the appropriate equations yields also the kinetic constants k_1 and k_{-1}[80]. This method is in principle rather simple and straightforward, however the situations to which it can be applied are rather rare but methods mentioned above - high pressure, elevated temperature, or a combination of denaturants or co-solvents, can shift the system to conditions allowing measurements by this method.

2.2.7 Residual dipolar couplings

The measurement of residual dipolar couplings (RDC) has been developed during recent years to a widely used technique supplying important long-range structural information for NMR structure determination studies[81, 82]. The application of RDCs to determine protein dynamics has a big advantage as it covers a broad range of time scale, spanning from sub-microsecond to sub-second events. The slower motional events can be separated from the fast ones by comparison with data derived from methods detecting fast motion (see e.g.[30]). This property makes RDCs in principle suitable to study protein folding and binding. There are, however, special difficulties considering technical aspects of the experiments as well as interpretation of the data. The proteins, which undergo (partial) unfolding, expose possibly hydrophobic parts, which can strongly interact with the alignment media. Additives like hexanol in PEG-derived alignment media can influence folding dramatically. From this aspect, stretched polyacrylamide gels represent the most generally suitable option[83, 84]. The molecules, which partly orient in magnetic field without any alignment media, form another favorable situation.

These difficulties limit the obtainable set of RDCs. Subsequently, the RDC values have to be evaluated to obtain dynamic information. Provided that the unfolded or intermediate states are populated significantly, averaging of RDC values can be observed. However, there still remains the issue of incorporating these data into structure calculations. “Time-dependent restrains” are in such cases employed in molecular dynamic calculations, which results in improved agreement of computed data with the experiment. This requires sophisticated calculations and will be mentioned in section 3.1.2.

2.2.8 RDCs of invisible states

In the majority of cases, the unfolded or intermediate state is so little populated, that a common measurement of RDCs does not result in meaningful information about it. This situation is similar to the one observed for chemical shifts, where the chemical shifts of lowly populated states are not visible in the spectra. The chemical shift information can be extracted using relaxation dispersion experiments (see section 2.2.4). The structural information derived from the chemical shifts is, in general, rather insufficient and is usually derived empirically. The RDC values, which provide structural information, give the possibility to develop a more robust method for the structure calculation. Very recently, it has been shown, that the RDC data can be also derived from relaxation dispersion measurements under weakly aligned conditions[85], provided that no problems occur with the selection of alignment conditions. The following discussion focuses on the most common measurement of RDCs from the backbone amides. The RDC values are usually extracted from the differences in line-splitting in isotropic and oriented media. Those differences correspond to the residual dipolar coupling (the isotropic value represents the scalar coupling). At the same time the splitting can be also derived from the difference in the chemical shifts. The fact, that the relaxation dispersion measurement allows extracting the chemical shift difference between the states under chemical exchange is used for the measurement of bond vector orientations of invisible states by RDCs.

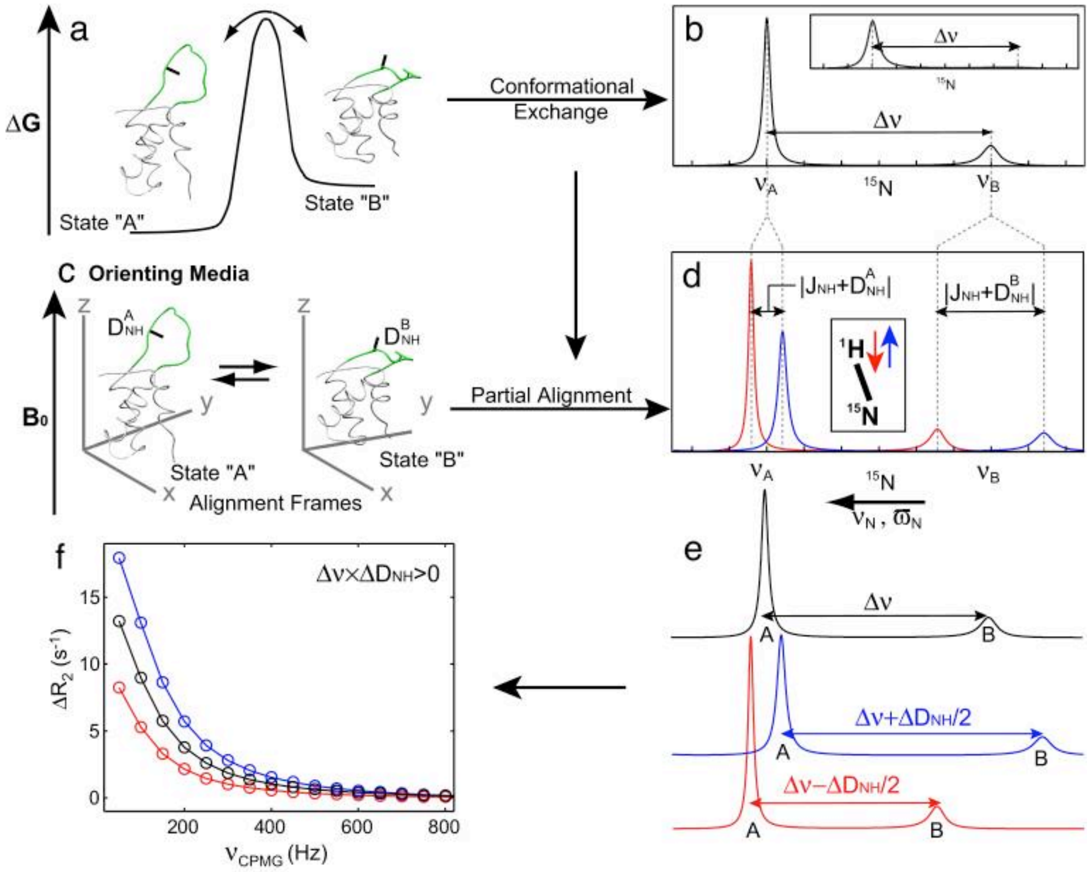


Figure 5: Measurement of amide bond vector orientation in invisible excited protein states. (a) Energy level diagram for a two-state exchanging system, where the loop (green) can exist in two conformations. (b) Resulting ^{1}H-decoupled ^{15}N spectrum for a single amide probe of conformational exchange between two states whose populations are highly skewed. In weakly aligning media (c) and without ^{1}H decoupling, each line is split by the sum of the ^{1}H-^{15}N dipolar and scalar couplings ($J_{NH}:93Hz$). Spectra resulting from the ^{1}H in the α- and β spin-states are shown in d in red and blue, respectively. (e and f) Separate ^{15}N CPMG relaxation dispersion experiments monitor exchange between ground and excited state conformations that are separated by $\Delta\nu$ (black), $\Delta\nu - 0.5\Delta D_{NH}$ (red), or $\Delta\nu + 0.5\Delta D$ (blue), from which ΔD_{NH} can be extracted. In f, intrinsic relaxation rates, $R_{2,\infty}$, have been subtracted from the dispersion profiles to emphasize their differences, $\Delta R_{2,\infty} = R_{2,eff} - R_{2,\infty}$. Reproduced from ref.[85].

The extraction of RDCs consists of 3 basic measurements (all of which are performed in the alignment media). The first measurement gives the chemical shift differences $\Delta\nu$ between "ground state" and "excited state" in similar way as described in section 2.2.4. Another two classes of experiments select the TROSY and anti-TROSY components in CPMG-based relaxation dispersion measurements[86]. The relaxation dispersion profile then sums up the contributions of chemical shift difference and residual dipolar coupling $\Delta\nu \pm \Delta D_{NH}/2$ (the scalar coupling is expected to remain constant between conformers, and thus does not contribute to the chemical shift difference). Additionally, to resolve the sign of RDCs, it is necessary to record the spectra at different static magnetic fields[87] (see also fig. 5).

The accuracy of the extracted RDC values presents the main difficulty and requires a statistical evaluation of the fitting procedure to judge whether the obtained values are sufficiently reliable. The method has been so far verified only on a single system. It offers, however, direct structural information about invisible states in form of bond vectors orientations.

2.2.9 Paramagnetic centers for structure and dynamics studies

A pair of spin labels incorporated by site-directed spin labeling (SDSL) allows the study of the distances and dynamics by Electron Paramagnetic Resonance (EPR)[89]. This gives rather limited information, since there is only one distance probe, which works between ca. 8-20 Å for a pair of the nitroxide SL. The distance range is increased to ca. 60 Å, if one of the spin dipole partners is a metal ion (e.g. Cu, Fe) in the metal cation-containing protein. Usage of this method makes it possible to obtain distance information in difficult cases like large membrane proteins and frozen solutions. The time-resolved spectroscopy can monitor the changes upon folding with millisecond resolution. The contribution of EPR spectroscopy to high-resolution structure studies is rather minor. EPR and Electron Double Resonance (ENDOR) spectroscopy is mainly used in studies on electronic structure of molecules, which naturally contain paramagnetic center.

The use of paramagnetic atoms or centers is also widely applied in NMR spectroscopy[90]. It allows to complement NOE data by long distance information, or to gain detailed structural information in systems, which are not suitable for measurement of (long-range) NOEs[88]. This is a common situation for partially unfolded proteins (see figure 6) and for large or membrane proteins.

The two main effects of paramagnetic labels are pseudocontact shifts (PCS) and paramagnetic relaxation enhancement (PRE). Both of them encode approximate distance and dynamics information. The drawback of SDSL is, that it requires a mutation and/or derivatization of the macromolecule. Commonly used residue for SDSL is a cysteine derivatized by a nitroxide radical (methanethiosulfonate spin label). The distance value can be obtained, provided that at least HSQC (TROSY) spectrum can be measured and assigned. The broadening (PRE) effect spans ca. 15 Å. The information content is limited by the fact that the distances are approximate and sparse. To obtain a more detailed picture, several mutations (several spin label sites) must be introduced.

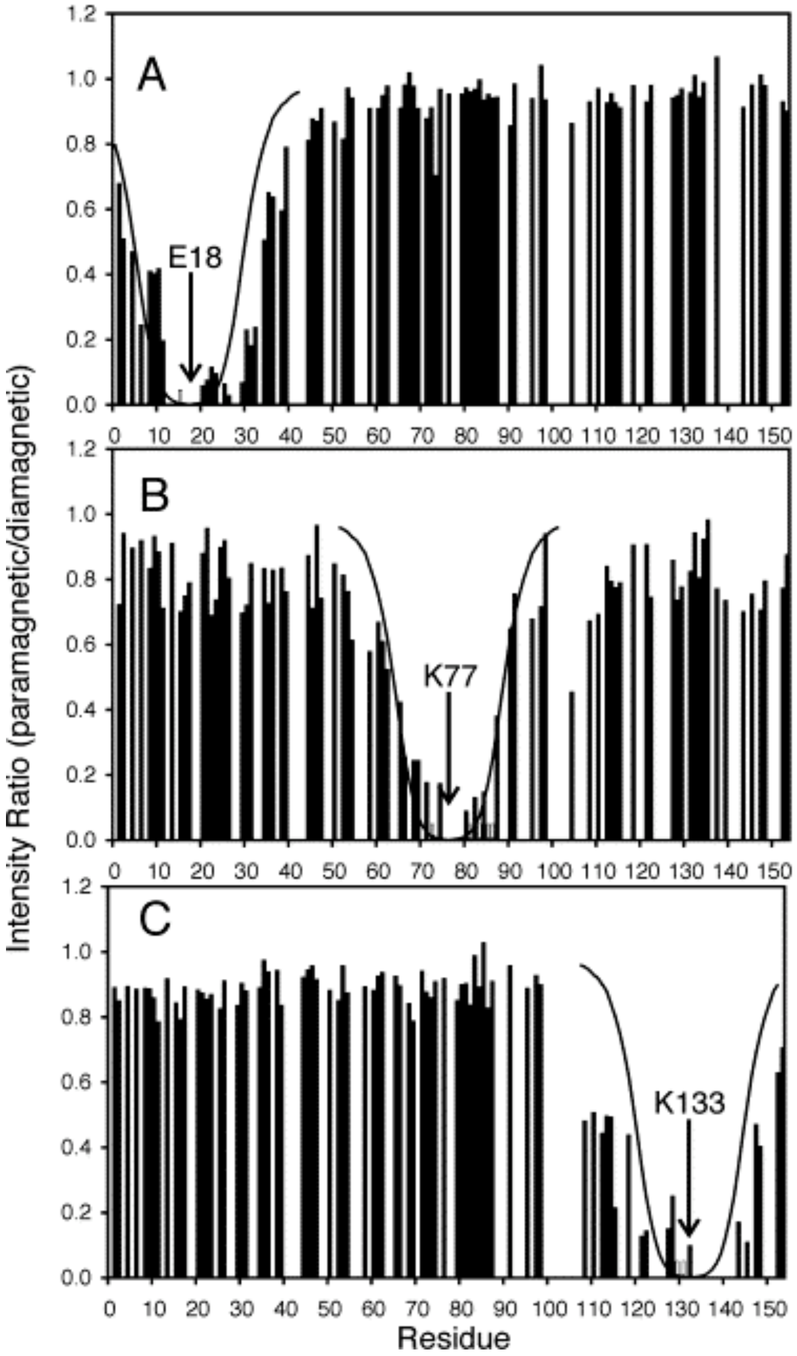


Figure 6: Paramagnetic enhancement to nuclear spin relaxation for unfolded apomyoglobin in 8 M urea at pH 2.3. The histograms show the experimental intensity ratios ($I = \frac{I_{para}}{I_{dia}}$) for each residue with an adequately resolved cross-peak in the ^1H–^{15}N HSQC spectrum of (a) E18C*, (b) K77C*, and (c) K133C*. Residues with HSQC signals broadened beyond detection are indicated by short open bars, while residues with insufficiently resolved HSQC signals are assigned ordinate values of zero. An intensity ratio of 1 indicates no effect of the spin label on an amide proton. An arrow indicates the location of the spin label. The superimposed curves show the simulated intensity ratios for a random coil polypeptide. The curves were simulated assuming r^{-6} broadening and a Gaussian distribution of distances between each residue i and the spin label. Reproduced from ref.[88].

One possibility to evaluate the spin-label data of the partially folded proteins is described below. If the polypeptide behaves as a random polymer, the paramagnetic broadening decrease with an increasing distance in the sequence from the spin-labeled residue. The expected broadening profile for the random coil then reflects the Gaussian distribution of distances between each residue and the spin label. Deviation from the random coil behavior indicates transient long-range contacts involving residues near the spin label. This method was extended and used in very recent work[91]. Here, the idea is that the PRE is observed in segments with above-average "values of the average area buried upon folding" (AABUF). Those are defined for each amino acid as the difference between the solvent exposed state and the mean solvent accessibility of the amino acid in the database of folded proteins. Clusters of residues with high values of AABUF indicate nuclei of local and also long-range contacts.

The commonly used nitroxide radicals do not cause sufficient pseudocontact shifts. From this point of view, very favorable properties were recently reported for the spin-label based on a classical shift reagent – an lanthanide ion - namely dysprosium ion (Dy^{3+}), chelated by 8-DOTA (1,4,7,10-tetraazacyclododecane-1,4,7,10-tetraacetic acid rigidified with 8 methyl groups). It causes large pseudocontact shifts, which depend on the orientation of the nucleus with respect to the DOTA spin-label. The magnetic susceptibility anisotropy also introduces measurable RDC values. Such a spin label then provides several valuable information sources leading to more detailed structural and dynamical knowledge[92].

3. Molecular modeling approach to protein folding studies

Without the help of computer modeling, generation of the structural data of (bio)molecules from the experimental data would not be practically possible. Whereas processing of the crystallographic and NMR data of rather rigid molecules is nowadays a routine procedure, evaluation of the data from more dynamic systems has many limitations. The same holds true for standalone molecular modeling (see e.g. ref.[93]). The dynamic events of protein folding are especially difficult to tackle for many reasons. It is, however, a way to obtain insight into the details of protein folding (see e.g.[94]), and can possibly help in designing experiments for validation.

Here are summarized the current possibilities in molecular modeling to help in experimental work, and to point out the possible pitfalls. The emphasis will be given to topics related to NMR spectroscopy, protein folding, protein structure prediction and advanced structure calculation.

3.1 Molecular modeling and experimental data

There are several ways in which computer modeling can help in the structural studies of macromolecules. i) First, probably the most common case is structure calculation where the experimental data form an input for the calculation of “static” structures. ii) A more advanced approach can also include variables, which are dynamically averaged. If this is done in an appropriate way, it gives a more accurate picture of the dynamics of the structure. iii) From the molecular dynamics simulation, performed with an appropriate accuracy without experimental restraints, many variables can be calculated and compared with the experimental data. This serves for validation or rejection of the proposed dynamical model. iv) There are examples of *de novo* structure calculation of small proteins without any direct experimental data.

Combinations of these kinds of molecular modeling are necessarily used for the protein folding studies. The first case has reached a certain maturity and will be not reviewed here (see e.g.[95-97]).

3.1.1 Comparison with experimental data

Almost all experimental data are both time averaged and ensemble averaged. From the molecular dynamic trajectory, it is in principle possible to calculate time averages that are equivalent to experimental data. If these agree with each other, the simulation may be considered to realistically represent the real system and may be used to calculate quantities, that are not experimentally observed, or that cannot even be observed. In this way, computer simulation complements the measurement (see e.g. ref. for recent review[98]). “If the molecular model and force field used in the simulation were infinitely accurate and the simulation could be made infinitely long, one would even not need to perform experiments”[99].

The fact, that the measured quantities are averaged over time and ensemble, encodes information about the dynamics of the system. The details are not commonly accessible, however. The experimental data does not contain enough information. To see the details, from which the experimentally accessible averages originate, it is possible to calculate many values from the molecular dynamic simulation. Such an approach is becoming widely used. In principle, any property, which depends primarily on coordinates and momenta (velocity), can be calculated and compared with its experimental counterpart. Experimental quantities commonly compared to molecular dynamics simulation are e.g. many NMR parameters (dipolar couplings related to NOE intensities, ^3J-coupling constants, RDC values, chemical shifts, relaxation times[100]), X-ray diffraction intensities, distances and other quantities derived from EPR and fluorescence spectroscopy. Also CD or AFM derived data and data from other techniques suitable for the study of the protein folding have been compared with molecular modeling data.

To illustrate how molecular modeling enables the proper interpretation of the experimental data, it is possible to choose any of the NMR observables. The most widely used property from NMR experiments are NOESY intensities, from which the interatomic distances are derived. The NOE peaks however originate from all of the conformers significantly populated in the sample. This results in the well-known problem, that only a subset of the observed NOEs can be fulfilled by the single static structure. The attempt to incorporate all of the NMR observables into a single structure results in a high target function during the process of structure calculation. The other effect is usually an unrealistically precise structure and/or a structure, which is actually never present in the sample and therefore wrong. The second well-known problem is associated with the scalar ^3J-coupling constants, which encode the dihedral angles through the Karplus relation[141]. In the majority of cases, treating ^3J-coupling constants as restraints in a static structure leads to many violations in the structure calculation. They are therefore rarely used in routine structure calculation, even though the information is in principle of similar value to the NOE crosspeaks (see e.g. ref.[101]). In favorable cases, both the NOE signals and ^3J-coupling values can be calculated from the molecular dynamic trajectory and compared with experimental values. The occurrence of NOEs, which were contradictory for a single static structure, as well as ^3J-coupling values, can then be described by time-weighted averages over the different conformers of the system. This can also identify errors in the original structure calculation, wrong signal assignments and also errors in or insufficient accuracy of the simulation (see also ref.[102]).

Before comparing simulated and measured data, it must be carefully considered whether a given property can be calculated for a given experimental and simulation setup. Otherwise wrong conclusions are likely to be made. In the following various problematic cases are discussed, which can occur when comparing values averaged over the conformational distribution $P(q)$ calculated from the simulation $Q_{ave,sim}$ with the experimental value $Q_{ave,exp}$:

- Q does not reflect the conformational distribution $P(q)$ since Q is not sensitive to a conformation. An example is the comparison of CD spectra of the beta-peptide studied in ref.[103]. CD spectroscopy generally contains little structural information, and a comparison with atomistic simulations is not appropriate. Already the interpretation of CD spectra in terms of distributions of secondary structure is relatively rough. Similar spectra can originate from very different conformations of the same peptide/protein as illustrated in the study of the above-mentioned beta-peptide. Atomic force microscopy is another technique, which does not provide sufficient atomic-resolution information, so that compensation of different sources of errors cannot be excluded. This possibly leads to wrong conclusions[104] and mis- or over-interpretation of the experimental data.

- Q does not reflect the $P(q)$ as Q is determined by rare conformations insufficiently sampled during the simulation. This case has also been observed during the comparison of calculated CD spectra with the experimental CD spectra of the beta-peptide in ref.[103].
- The simulation is insufficiently accurate, which can be due to: i) the omission of relevant degrees of freedom - for example - when the continuum solvent approximation is used, ii) the force field not being sufficiently accurate regarding the used relations and/or constants. iii) the inaccuracy of the numerical integration of the equation of motion, iv) the use of inappropriate thermodynamic boundary conditions, e.g. the type and strength of a temperature and pressure coupling. Also spatial boundary conditions can be set incorrectly, e.g. in the use of periodic boundary conditions together with a corresponding treatment of long-range interactions. The resulting artificial periodicity causes mutual influence of the mirror images.
- The experimental data are inaccurate or wrong.
- The compared quantities are related, but not the same, or quantities were derived based on simplifying assumptions/relations, for example, comparing atom-positional fluctuations with crystallographic B factors or ^3J-coupling constants with the dihedral angles.
- Simulation and experiment were performed under different conditions e.g. in a crystalline and solution state, at different temperatures, under different pressure, pH etc.

A particular problem for folding studies is the time scale, which the simulation can span. Only small systems with fast folding events can be appropriately averaged over the μs range. The comparison of measured and calculated folding rates remains problematic in any case. Reversible *in silico* folding was reported only in the late nineties[105, 106]. The possible methods addressing this problem are discussed in section 3.2.

3.1.2 Oscillating time-dependent restraints and local elevation for dynamic restraints

Time-dependent restraints are used to generate an ensemble of molecular conformations that are in agreement with the measured ensemble averages for particular observables. This is particularly relevant for highly dynamic structures connected with protein folding and binding.

Since observables as NOEs or ^3J-coupling constants generally depend in a non-linear manner on the 3-dimensional structure of a biomolecule, it is not straightforward or even impossible to derive the probability distribution of molecular conformations from such experiments. Particular issues are the multiple-valuedness and the sizeable gradients of the Karplus curve, which relates the ^3J-coupling constants measured in NMR experiments to the corresponding dihedral-angle values. This causes severe problems when trying to restrain the dihedral angle that corresponds to values close to the extrema of the Karplus curve. The introduction of a factor, oscillating with time into the restraining penalty function, alleviates this problem and enhances the restrained conformational sampling[99].

Another method is based on “adaptive restraints”[107]. It uses a local elevation search[108], in which the potentials are increased for those conformations that were already visited during the simulation. This is done in a slow (adaptive) way, until either the instantaneous or the average value of the coupling constant fulfills the experimental observation. The time average of the ${}^{3}J$-coupling is calculated during the MD simulation, and if this does not match with the experimental value, the dihedral angle is effectively forced to adopt a range of values different from those sampled before. The method overcomes the described difficulties and makes the ${}^{3}J$-values valuable for structure calculations.

These types of refinement generally rely on employing a high quality force-field, because they introduce more conformational freedom to the molecule. These techniques present not only a successful method for dynamic restraining of dihedral angles based on ${}^{3}J$-values, but also a guideline for utilizing other experimental data on dynamic systems, namely RDCs.

3.1.3 Molecular dynamics in water/methanol mixture and at elevated temperatures

In the context of the current work, it is of particular interest to study the effect of methanol and elevated temperature on the conformational behavior of proteins and peptides. The insight from molecular simulations and from a comparison with the experimental data is briefly introduced using the example of ubiquitin. A second example is a thermodynamic study of a beta-heptapeptide belonging to the so-called class of foldamers.

Ubiquitin is one of the experimentally best characterized proteins. A series of molecular dynamic simulations under different conditions was used to shed light on some of the effects important for stability of the structure[109]. The simulation shows the possibility of relating changes of thermodynamic variables to changes of structure and dynamics of the protein-solvent complex. As an example it was shown that the heat capacity change upon folding may be attributed to the solvent (water) ordering effect of exposed hydrophobic residues upon unfolding. This effect is absent in 40% methanol. At room temperature, methanol induces stabilization of secondary structure and reduction of tertiary interactions. Overall, methanol slightly destabilizes the protein at neutral pH and slightly stabilizes it at low pH. This is consistent with the expected ability of methanol to moderate hydrophobic interactions in a nonspecific way. Solvation of exposed (hydrophobic) core residues almost exclusively involves methanol molecules, despite their larger size compared to the water molecules. This is probably related to the higher hydrophobicity of this solvent. The protein-solvent hydrogen bonds are almost equally distributed between water and methanol, despite the smaller total number of methanol molecules in solution (40% in this study) and the fact that methanol can only donate a single hydrogen bond (as opposed to two for water). There is clearly preferential solvation of the protein by methanol in the mixture, that is, the solvent composition in the immediate vicinity of the protein is richer in methanol compared to the bulk solvent composition.

Peptide folding entropies and free enthalpies were calculated in an illustrative study of a simple beta-heptapeptide in methanol[110]. The simulation also allowed for the estimation of the temperature dependency of thermodynamic parameters like enthalpy, entropy, heat capacity and free energy describing the folding/unfolding equilibrium of the peptide.

These studies demonstrated, that the solvent and temperature effect on macroscopic thermodynamic properties can be, with the current classical force fields, related to the microscopic properties of the system.

3.2 Ways to enhance sampling

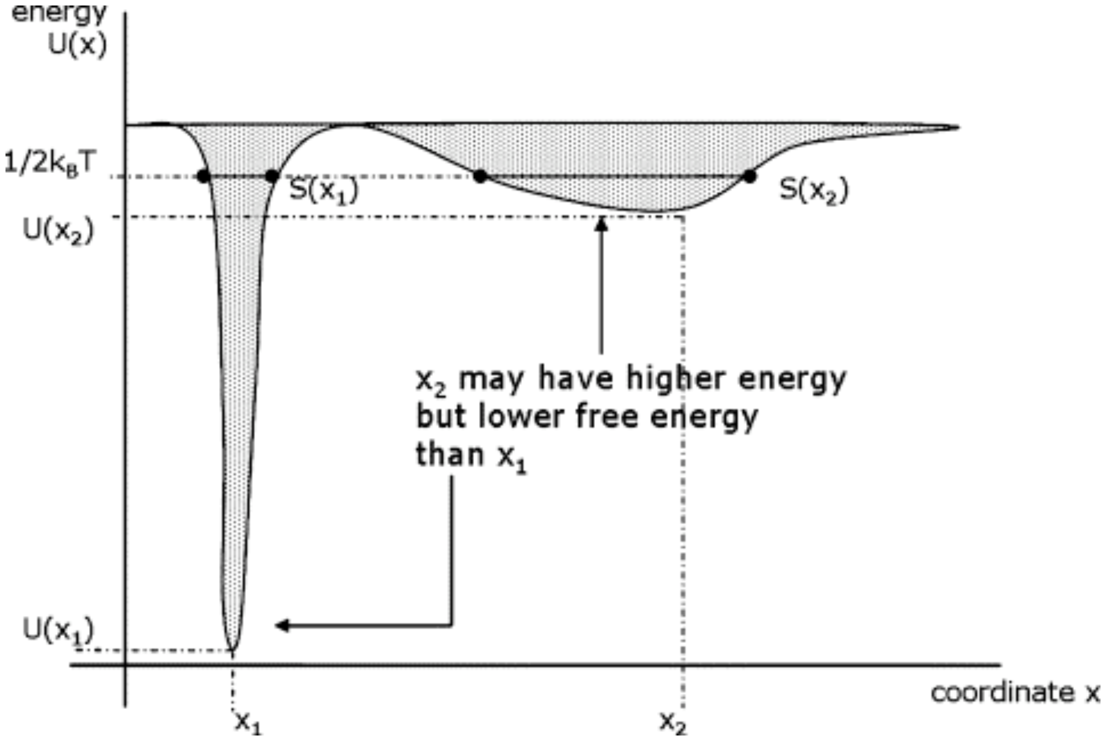


Figure 7: Energy (U)–entropy (S) compensation at finite temperature T. Reproduced from ref.[111]

As outlined in the previous section, the critical problem in molecular modeling, when applied to protein folding, is the time scale which it can span. The interactions in the molecular systems are by nature quantum mechanical. Fortunately, classical molecular dynamics on an atomic level at room temperature can approximate many dynamical events and interactions in biomolecules with sufficient accuracy allowing to study systems under realistic conditions. But even with this approximation, only those events, which occur within μ s can be observed. This is the reason, why *in silico* folding for moderate-sized proteins cannot be observed, even if the underlying physical description is correct. Especially for structure calculations of biomolecules based on experimental data, more efficient sampling techniques were devised (see e.g.[111] for a recent review on sampling methods). The most commonly used method, as applied to NMR data, is Simulated Temperature Annealing[112]. To cover the processes like folding transition states for which no or only very limited experimental data exist, is even more difficult. The enhanced sampling methods will be systematically covered in the following part. Before that, conformational sampling methods will be briefly classified according to properties of the conformational set they produce. They can be divided into three types:

1. Methods that generate a set of non-related, low-energy configurations. This is the most frequently used type for structure calculations from NMR data. The

generated “bundle” of structures should represent the set, which optimally reflects or fulfills the experimental data. The deviation among these structures shows the uncertainty of the structure definition, which is usually expressed as the RMSD. The fact that the different structure calculation protocols produce slightly different set of structures indicates that the RMSD and the concept of a bundle of structures contains only orientational information. See figure 7 for an illustration of the difference between a low-energy structure, e.g. a structure with little residual violations of restraints or force-field parameters, and a structure with a low *free* energy.

2. Methods that generate a properly (Boltzmann) weighted set of configurations. A typical example is the Metropolis Monte Carlo simulation. The thermodynamic properties can be calculated from the sampled configurations. An example of its use is the calculation of the free energy of binding or complexation processes.
3. Methods, that generate a dynamics trajectory of properly (Boltzmann) weighted configurations. An example is classical MD, which is mentioned repeatedly throughout the introduction as a basic molecular modeling tool.

3.2.1 Possibilities for speeding up sampling

The total amount of computation time can be reduced by the introduction of cut-offs for non-bonded interactions. This is easily done for short range VdW interactions. Various methods are available for long-range electrostatic interactions, e.g. by Ewald summation[113]. Another common alternative is to approximate the medium beyond the given cut-off distance from a specific atom by a dielectric continuum with corresponding permittivity and ionic strength (see e.g. ref.[114] - relevant for protein folding).

Another approach is coarse-graining of the system. Using the all-atom classical mechanic approximation is already coarse-graining with respect to the more appropriate quantum mechanical description. The next widely used level is coarse-graining of the aliphatic CH_x groups leading to aliphatic united atom force field as done in the GROMOS force field[115]. The coarse-graining can continue to amino acid level, which is used in the low-resolution step of the Rosetta structure prediction program (see sec. 3.3). From the biologically relevant applications, there are known examples of coarse-graining of acyl chains in membrane simulations. Coarse-graining can speed up the calculations many-fold, and if well chosen, can preserve the precision required for a given purpose.

3.2.2 Heuristic methods

The heuristic method is a general designation of the search methods, which visit only a tiny fraction of the available configuration space. This aims to be a representative (in a Boltzmann sense) set of the system configurations. Examples are given below:

Soft core atoms

The soft core atoms technique is a conceptually very simple heuristic method. It alleviates the problem of many energy barriers arising from overlap between atoms when using the usual hard core repulsive potential. Softening the strong and steep repulsive potential lowers these barriers, smoothens the potential energy surface and leads to strongly enhanced sampling[108]. The concept has already been used for protein folding in the eighties[116].

Biasing the search

Biased sampling along the pathway can enable access to processes in a biomolecular system, that occur on a time scale far beyond those accessible by a standard MD. There exists no real solution for processes that are inherently slow, because extensive sampling connected with the complicated reorganization is necessary. However, if the processes are infrequent but fast when occurring, there exist the possibility to enhance the occurrence of these rare processes. The rareness is generally caused by high-energy barriers (transition-states) separating the states. The oldest approach to sample transitions is to define a reaction coordinate, or in other words a transition pathway, and sample along this path using a biased potential energy term or to use umbrella sampling[117]. This technique, together with so-called targeted MD[118] is well suited for simple reaction pathways, for which the transition-state is known.

Transition path sampling

A more sophisticated methodology is the Transition Path Sampling (TPS) (see e.g. ref.[120, 121]), which finds transition pathways for infrequent events and does not require knowledge about the transition mechanism or the transition state. Only the end states A and B have to be defined. The transition pathways are extracted from the simulation results. From systems relevant to protein folding, the method was tested and analyzed on an alanine dipeptide and a 16 amino acid peptide. Even though the method is able to properly sample the possible transition pathways of rather complicated systems, it remains applicable only to simple examples of protein folding.

The states A and B are stable in the sense that trajectories will stay in that state for a time long on the molecular time scale (e.g. vibrations). The state has also defined regions (see fig. 8). The procedure starts by finding the initial trial path between the two states. This can be done for a folding process by simulating the originally folded peptide at elevated temperature to obtain a possible unfolding pathway as e.g. in ref.[122]. The path is divided into time points. From these points, new simulations are started with random initial settings. The original implementation[123], which employs Monte Carlo dynamics, leads to divergence of the original pathway in a random direction. It was, however, found, that the procedure fails for the “diffusive” barriers of protein folding. Starting from the time point near A, the simulation reaches state B with a negligible probability. The deterministic MD on the other hand does not diverge sufficiently from the original pathway. Improved capabilities arise by the use of the Andersen thermostat[124] as a random noise generator for MD, this increases the sampling efficiency[125].

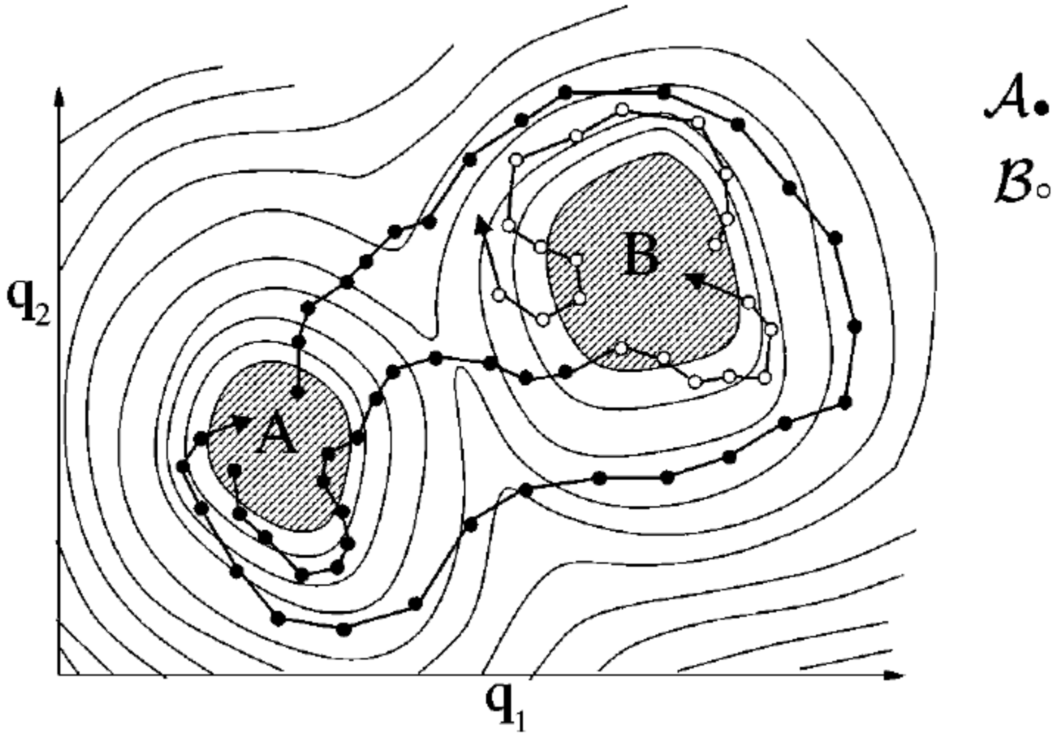


Figure 8: Example of phase space regions in TIS. Thin solid curves denote the free energy contour lines. q_1 and q_2 are two arbitrary projections of the degrees of freedom. A and B are the two stable states. The dots on the three shown trajectories indicate the positions of the system at successive time steps. The overall state A and B are indicated by black and white dots, respectively. Only one trajectory starts in A and ends in B and is therefore a true reactive event. The system changes from state A into B into when it enters region B for the first time. It can leave stable region B shortly thereafter, but never go back to A in a short time. The stable regions have to be chosen to fulfill that condition. Reproduced from ref.[119].

The algorithms to construct the new path can be described in the following steps: Start with an initial path of L time points connecting A and B. Let t_A be the time just before leaving A and t_B the first time the system enters B on the old path. Now choose a point on the barrier between t_A and t_B with a uniform probability (or with a biasing function). Next, integrate the equations of motion forward to t = T and backward until time t = 0 using the Andersen thermostat as a stochastic noise generator. If the new trial path between t = 0 and t = T still connects A with B, then this walk ("shot") from the time point is accepted with a probability, which takes into account (time) distances between states A and B on the old and new path. If the backward path reaches A but the forward shot fails to end in B, the forward trajectory is rejected. The old backward path from t = t to 0, however, is replaced by the new backward trajectory, if it is accepted, based on the calculated probability. Similarly, if the forward path reaches B but the backward shot fails to end in A, the forward trajectory is substituted from t = t to T and accepted with the analogical acceptance probability.

Method TPS was successfully applied on the extensively studied model folding system of the 16-residue C-terminal fragment of the protein G-B1[126]. This peptide forms a beta-hairpin structure, which is stable at room temperature molecular dynamic. The unfolding can be seen upon simulations at elevated temperature (400K). Computation of several hundreds of thousands of paths was a time consuming task (3 months on 8-processor PC cluster). It was then possible to calculate the rate constant in agreement with the experiment. The method is therefore capable not only to find possible transition pathways, but also of characterizing the kinetics of the event. The method Transition Interface Sampling (TIS) is a more recent improvement of the TPS method[119, 127].

As mentioned in section 3.1.1, calculation of a rate constant would not be possible without the use of the heuristic method. Even for the rapidly folding systems (as in ref.[105, 106]), for which the repeated folding and unfolding can be observed *in silico*, extraction of the rate constant is not obtainable using common MD simulation.

Minimum energy path

The Nudged Elastic Band method is another method to determine the transition pathway[128]. An example of application relevant for protein folding is a conversion of an alanine decapeptide α-helix to a π-helix[118]. This is relatively small rearrangement that comprises breakage of the hydrogen bonds and slight dihedral angle changes. To use this method, it is necessary to generate a set of intermediate structures, which connect the limit states. For this system, 51 intermediates were prepared. These states are then connected by “elastic bands” with a defined force constant. The transition is in this way targeted and simultaneously sampled by MD and energy minimization. This technique supplies “only” the minimum energy pathway, and needs a clearly defined predefined way, which limits the applicability. Nevertheless, it is one of the methods, which can be used in a real situation and compared to the experimental data.

3.3 Rosetta, *de novo* structure prediction

The detailed understanding of the protein folding can, in principle, lead to prediction of structure from the amino acid sequence. As a validation parameter that reports on how the predicted structure corresponds to the native structure should naturally be the free energy. The difficulty to calculate the free energy can be seen from a calculation of the heptapeptide already mentioned in section 3.1.3. However the current available computational power compelled substantial simplification in this assessment, it led to encouraging results. From the desired dynamics information, as stressed throughout this work, it is necessary to deal with much simpler data, namely with static structures, and assign single energy values to these structures, which will be used for ranking. This corresponds to generating arbitrary, low-energy structures (case 3 of section 3.2). One problem is the assignment of the lowest energy value to the conformation which is closest the native structure. The second crucial difficulty is finding an efficient method to generate these conformations. Some approaches for the enhanced sampling were mentioned in section 3.2. For *de novo* structure prediction, combination of heuristic techniques results in a viable algorithm. Such an algorithm, incorporated in the program Rosetta, will be briefly described bellow[129, 130].

The force field, which evaluates the energy of the input structure, is based on the short-range interactions - primarily van der Waals forces, hydrogen bonding and favorable desolvation, while it neglects long-range electrostatic interactions. This obviously limits the type of correctly predicted proteins to those with tightly packed hydrophobic cores, which are further stabilized by hydrogen bonds. Highly dynamic sites, as well as sites, where the electrostatic and pivotal solvent interactions are important, are evaluated with less accuracy. These are present e.g. in the functional sites of enzymes. The subtleties of the specific prosthetic groups are also not captured.

The conformational sampling protocol is even a more difficult task. To make this task practically tractable, it is necessary to divide the conformational sampling into a local, high (atomic) resolution part, and a more global, low resolution sampling. The latter one immediately looses the details necessary for a rational *ab initio* energy assessment. The low resolution step therefore generates a large number of low resolution folds based on statistical data. This comprises a search of sequence homologs and a retrieval of the structure fragments from the PDB database. Similarly, low resolution structures produced by experimental methods can be used in this low resolution step. The NMR experiments in less favorable cases supply only sparse information, which can be then used as an input, the same is true for X-ray diffraction data. Approximate sparse distances are derived from the NMR experiments with paramagnetic labels or from EPR spectroscopy (as sketched in section 2.2.9). The only requirement for choosing the method is that it has to generate a set of structures, which contains the structure within the “radius of convergence” of the subsequent high resolution refinement step.

In contrast to the low resolution step, the high resolution step proved to be able to lead the initial input to a native-like structure without any statistical or experimental input, which is described in more detail in the following section.

3.3.1 Rebuilding and refinement protocol

The high resolution step consists of several steps. The first stage is an energy optimization of the input ensemble, which combines Monte Carlo minimization with remodeling of the side chains. The next step identifies regions of high variability. These regions are then preferentially rebuilt, followed by another minimization comparable to first step, after which the final result emerges. If the resulting set of structures did not converge to a unique target, it is then used as a seed in a evolutionary algorithm and subjected to the next round of the rebuilding and refinement protocol.

This protocol is described in more detail to illustrate the complex heuristic method which utilizes the modular buildup of the protein[131].The simplified description of the protocol follows these steps:

- Input coordinates set from NMR, CM or DN

This is the input coordinate set from either experimental method, comparative modeling or *de novo* modeling.

1. All atom optimization

This step refines the initial ensemble, it combines Monte Carlo minimization with remodeling of the side chains to optimize side chain packing and hydrogen bonding according to all atom force field. Briefly, in each complex Monte Carlo step, a random perturbation to the protein backbone torsion angles is followed by discrete optimization of the side chain conformations, which allows efficient crossing of side chains torsion barriers. Then, an optimization of the side chains and backbone torsion angles is carried out before deciding on whether to accept this complex Monte Carlo step. The final minimization means, that each point on the landscape is mapped to the closest local minimum. Although it is possible to recognize near-native conformations based on their lowest energies, this all-atom refinement alone does not consistently produce significant improvements in model quality.

2. Identification of variable regions

The second step in the protocol is the identification of variable regions in the ensemble of refined models. It was observed that a significant correlation between the extent of variation in coordinates of an amino acid residue in the refined structures and the deviation of the coordinates of the residue in the refined models from the native structure exists. The positions exhibiting small variance across the models are usually quite close to the correct structure, whereas positions for which the variance is large often deviate considerably from the native structure. This correlation arises from the relatively short range of the force field and the energy gap between the native structure and the models. Because the energy of the entire system is roughly equal to the sum of its parts, for most portions of the protein, the correct conformation will be lower in energy than non-native conformations. Regions of the protein that reproduce the native conformation are likely to converge on this conformation and thus exhibit less variation, whereas locally incorrect conformations are likely to be spread throughout the landscape and to exhibit more variation.

3. Stochastic rebuilding

The third step in the protocol targets extensive sampling to those regions most likely to be in error. A fragment-based segment rebuilding method is used to completely rebuild regions of models with relatively high variations in the model population.

The coordinates of randomly chosen nine, three or one residue units of the identified region are generated and (re)inserted into the structure, so that the backbone chain is closed again[132]. The number of trials is 10 times the number of rebuilding residues. A Metropolis Monte Carlo criterion is used to accept or reject the newly inserted fragment based on a low-resolution energy function. In the second half of 10 rounds of rebuilding, the backbone torsion angle moves are performed together with the residue insertions. If after these 10 rebuilding stages the chain break between the rebuilt part and the rest of the molecule remains larger than 0.2 Å, the rebuilt region is expanded by one residue on both sides and the chain-break closing procedure is repeated along with the flexible variation of the starting values of the torsion angles. The regions are allowed to expand up to 5 residues on both sides in the stage of closing the backbone chain.

Because the regions that are incorrect cannot be identified unambiguously, many independent calculations are carried out, in which different segments in the higher variation regions are randomly selected for a complete rebuilt.

4. All atom refinement

The partially rebuilt models are then subjected to the Rosetta all-atom refinement protocol. In the segment rebuilding process, side chains are initially represented as soft interaction centers and the connectivity of the chain is temporarily broken, thus permitting the traversal of much larger barriers than those crossed through an all-atom refinement alone. The torsional states of backbone and side chains are evaluated using knowledge-based potentials derived from amino-acid-specific Ramachandran maps, the rotamer probabilities and angle standard deviations in the rotamer library developed by Dunbrack and Cohen[133]. A single movement in the all-atom refinement protocol consists of the following steps. i) The backbone torsion angles are perturbed in one of several ways. This can be small random (up to 1.5 degrees) movements or concerted movements of ϕ and ψ angles in opposite direction (up to 3 degrees). Larger rearrangements cause movements of torsion angles in 1-3 residues, which follow the minimized region. Small-wobble movements involve 10-20 degrees random change in a backbone torsion angle in a single residue, which is then followed by a minimization including the 3 adjacent residues. ii) In the next step, the side chain rotamer conformation for the new backbone conformation is optimized. iii) An energy minimization with respect to either the backbone degrees of freedom (first half of the refinement procedure), or backbone and side chain degrees of freedom (second half of refinement procedure) is performed. iv) The compound movement (steps i-iii) is accepted or rejected according to the Metropolis Monte Carlo criterion. These combined movements extend the Monte Carlo minimization procedure found to be quite powerful in previous studies[134] by incorporating a discrete optimization of side chain conformations. This allows energy-directed barrier hopping at the level of side chains.

Finally, small scale refinement is done. After this procedure, a set of low-energy models is obtained.

- If the lowest energy refined structures did not converge to a unique structure, an evolutionary algorithm follows. It includes clustering of the set of 200 models and independent rebuild/optimization of each of the cluster representatives. The results, which started from distinct conformers, are then compared. This evolutionary strategy combines a search in the low energy region of the potential energy landscape together with maintenance of a search broad enough to avoid convergence to the next local energy minimum.

It was recognized, that there is an apparent need for a contribution of entropy by dynamics during the evaluation of the structure quality. The second problem in the current setup is the limited sampling. Even the described protocol does not sample conformational space thoroughly enough to find the global minimum, particularly if the input low-resolution structures are far from the correct one.

The protocol was also successfully applied for improving conventional NMR structures. These may provide valuable solutions to the X-ray crystallographic phase problem in molecular replacement trials[131].

3.3.2 *De novo* structure prediction of membrane proteins

Membrane proteins represent a specific challenge for the folding prediction attempts. Their folding *in vivo* involves insertion into a membrane, assembling and reorientation of the trans-membrane (TM) segments leading to final tertiary and possibly also quaternary structures[135, 136]. Recently, the above-described protocol was extended for membrane proteins including continuum solvent models. The bulk water, headgroup membrane region and hydrophobic core are modeled this way. The force field is similar to the one used for soluble proteins. It explicitly includes hydrogen bonds together with their orientation dependence and bifurcation. Beside that, it also includes weak hydrogen bonds between C-H and O. All these details proved to be important for prediction of the proper side chain conformation. The successful examples include prediction of an interface-bound peptide, docking of individual TM helices using fixed backbone and flexible side chains, and *de novo* structure prediction with a RMSD < 2.5 Å from the native structure for proteins < 150 amino acids. For the all-atom refinement protocol, it is crucial to supply near-native models with RMSD < 4 Å from the unknown native structure[137].

3.3.3 *De novo* design of enzymes

Recently, advances in protein structure prediction were manifested by computational design of an artificial enzyme processing a non-natural substrate[138]. It comprises design of possible geometries of the catalytic sites, which are simultaneously compatible with several transition states of the multiple-step reaction pathway. Scaffolds that allow the construction of the desired active site were found. Further selections were done for binding affinity of the transition state structures in the active site pocket. The total number of 72 designed enzymes with different protein scaffolds was recombinantly prepared yielding 70 soluble proteins. 32 proteins, spanning various folds, have detectable enzymatic activity, however usually worse than the naturally occurring enzymes. Another example of successful use of this approach is the re-design of the DNA endonuclease, which resulted in altered target site specificity[139].

4. Scope of the work

Cell-cell recognition is an event of prime importance in a variety of biological phenomena occurring especially in metazoa. Cell migration, organ formation, immune defense, or microbial infection are just several examples with complex underlying mechanism. The specificity of the interaction is mediated by various tissue-dependent receptors. Since the carbohydrates largely cover the cell surfaces, they often participate in specific interactions occurring on the cell surfaces. In addition, the extracellular matrix, which forms a substantial part of our body, is also largely built out of carbohydrates. The chemical constituents of outer membranes from mammalian cells and bacteria are significantly different. The main matrix of mammalian outer membranes is formed by phospholipids, whereas additional coating structures are present covering these membranes in bacteria. In Gram-positive bacteria a thick peptidoglycan layer is built around the phospholipid bilayer. In Gram-negative bacteria, the peptidoglycan structure is much thinner and coated by an additional phospholipid-containing bilayer, whose outer leaflet is mainly composed of lipopolysaccharides (LPS). Our work has focused on two examples of specific recognition between surface carbohydrates and peptide/protein (lectin) binder, since only little structure information about the interaction of these oligosaccharides with proteins is available. Furthermore, not much is known about the influence of the membrane environment.

In one study, we presented a synthetic glycolipid system that can be readily incorporated into phospholipid micelles and that allows the study of cell-surface-exposed carbohydrate units by high-resolution NMR techniques. We showed an efficient route for the synthesis of glycolipid compounds that contain mannose, mannobiose, or mannotriose coupled either directly to an alkyl chain or through a poly(ethylene glycol) linker. Furthermore, we have validated our model system by measuring the binding of cyanovirin N (CV-N), a cyanobacterial protein that binds with nanomolar affinity the terminal arms of high-mannose structures of the HIV surface-envelope glycoprotein gp120. From the results of chemical-shift mapping with uniformly ^{15}N-labelled CV-N, we made the important conclusion, that binding of high-mannose structures anchored by aliphatic chains onto the DPC micelle with CV-N occurs at sites similar to those involved in binding the non-conjugated carbohydrates. We characterized the insertion of the glycolipids into dodecylphosphocholine (DPC) micelles by measuring translational diffusion, and we observed that the diffusion constants of the glycolipids were very similar to those of the DPC micelles themselves, but significantly deviated from those of the free carbohydrates. We also present an experimental proof that the glycolipids remain inserted in the micelles while binding to CV-N. Finally, by addition of a ligand that had a higher affinity to CV-N but which was not attached to a lipid anchor, CV-N could be released from the glycolipid and, hence, from the micelle-associated state. Besides that, the studies revealed that protein-carbohydrate association in our system was not disrupted by the presence of DPC as a membrane mimetic.

The second study was devoted to the interaction of lipopolysaccharide from Gram-negative bacteria with polymyxins which are lipopeptide antibiotics. Antimicrobial peptides against Gram-negative bacteria can interfere with the integrity of the LPS layer. Even if no specific interaction with LPS exists, they need to pass the LPS layer in order to gain access to the cytoplasmatic membrane.

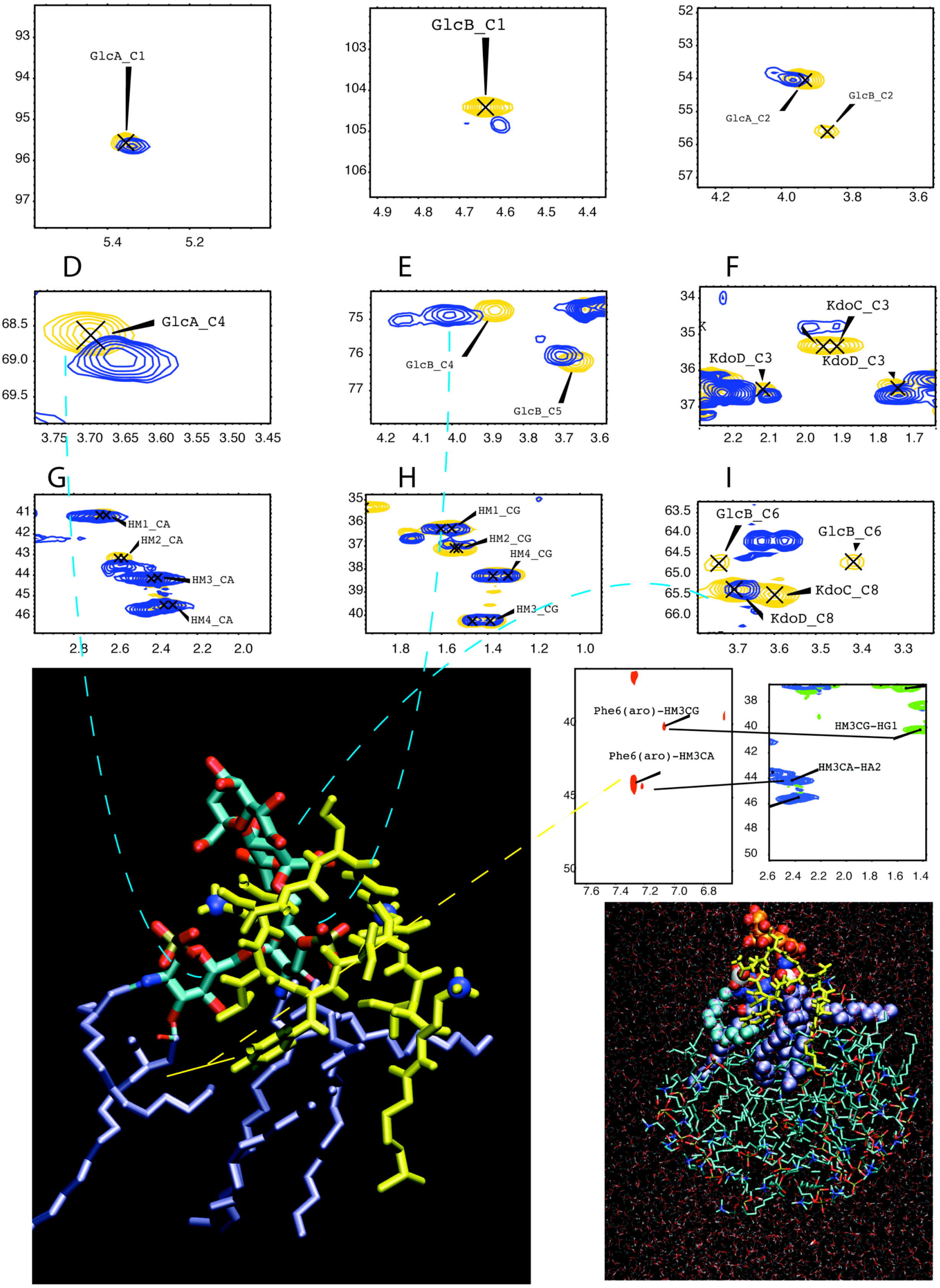


Figure 9: Scheme showing the usage of NOE data (right down) and chemical shift mapping data (upper nine spectra extensions) to calculate the model of polymyxin B-LPS complex. On the left side the structure of the polymyxin B-LPS complex is shown without explicit solvent. Carbohydrate moieties of LPS are colored according to atom types, PMX-B is shown in yellow with amine nitrogens emphasized as blue balls. The structure model on the right includes also DPC molecules and water molecules (colored according to atom type). PMX-B is depicted as yellow lines, and LPS as van der Waals spheres.

Cationic lipopeptide antibiotics like polymyxins, characterized by a heptapeptide ring and a fatty acid tail, are mainly used for topical treatment of infections caused by Gram-negative bacteria. We presented a study of the interactions of PMX-B, -E and -M with LPS from a deep-rough mutant strain of *E. coli*. First, a method for efficient purification of biosynthetically produced LPS using RP-HPLC in combination with ternary solvent mixtures was developed. Then LPS was incorporated into a DPC micellar membrane model and its interaction with polymyxins was studied by heteronuclear NMR spectroscopy. Data from chemical shift mapping using isotope labeled LPS or labeled polymyxin, as well as from isotope-filtered NOESY experiments, shows the mode of interaction of LPS with polymyxins. Using simulated annealing protocol and restraints derived from intermolecular NOEs, we have obtained a model of LPS-PMX-B complex, which was also in accordance with chemical shift mapping data. The model was further refined by MD calculations, which included DPC micelles and explicit water. In the modeled complex the macrocycle of PMX-B is centered around the phosphate group at GlcN-B, and additional contacts from polar sidechains are formed to GlcN-A and Kdo-C, while hydrophobic sidechains target the acyl chain region.

Helical hairpin folding

In this part of the work, we have probed the formation of the hairpin structure between α-helix and type-II polyproline helix in the peptide PYY. The peptide PYY (pPYY) represents a unique protein/peptide model system. It is a small 36 amino acid residue, C-terminally amidated peptide, easily handled and measured by high-resolution NMR. At the same time, it possesses a complex variety of features, allowing the study of a range of dynamic events and peptide folding. Its structural and dynamical properties readily respond to moderate tuning of the environment, mutagenesis or the presence of a membrane mimic. Its tertiary structure - the α-helical hairpin is unusual for such a small peptide. The structure was described by our group and present in ref.[140] . It also forms a quaternary structure, which consists of the weekly dimerizing tertiary structure.

Structure of pPYY

The pPYY structure in water was re-determined in this work based on an extended set of NOE and RDC data. The data shows typical features of a more dynamical structure (as mentioned in section 3.1). In particular, there are more NOE contacts than expected for a single structure. Incorporation of all of those NOEs leads to an unrealistically well-defined structure, inconsistent with the dynamics measurements. Many weak NOEs and also RDCs originating from the flexible part of the molecule were therefore removed from the structure calculation. The resulting structure presents an optimized approximation available by standard methods.

Recently, π-Pro interactions were postulated to contribute substantially to the stability of long-range contacts. In this study we replaced Pro residues 2, 5 and 8 as well as Tyr 20, 21 and 27 by Ala to produce all single mutations as well as the quadruple Ala 2,5,8,14-PYY mutant by site-directed mutagenesis. The destabilization of the helical hairpin in these mutants was probed by measuring changes in internal backbone dynamics by means of the heteronuclear NOE between the amide nitrogen and its directly attached proton. The $^{15}N\{^{1}H\}$-NOE data indicated that none of the Pro mutants were stably back-folded. For the quadruple Pro mutant the $^{3}J(HN\alpha)$ scalar couplings are > 6 Hz for all residues in the N-terminal segment, which is an indication of conformational averaging. By contrast, for the single Pro mutants the N-terminal segment was not fully flexible. The rigidity of the C-terminal α-helix is highly correlated to the rigidity of the turn and the adjacent amino acids. Moreover, the lack of long-range contacts destabilizes in particular the last two turns of the helix.

The structure calculation indicated possible presence of hydrogen bonds of the hydroxyl group of Ser13 with the amide proton of Glu15 and to the side-chain carboxyl group of the latter residue. We conclude that the presence of a hydrogen bond donor at position 13 is crucial for correct formation of the beta-turn. The shift of Pro from position 14 in PYY to position 13 in another peptide from the neuropeptide family, or the replacement of Ser13 by Ala is sufficient to disrupt the tertiary contact. The results showed that helical hairpin formation in this peptide was both reversible and cooperative and the specific hydrophobic contacts between the polyproline and the α-helix provide the main driving force for folding.

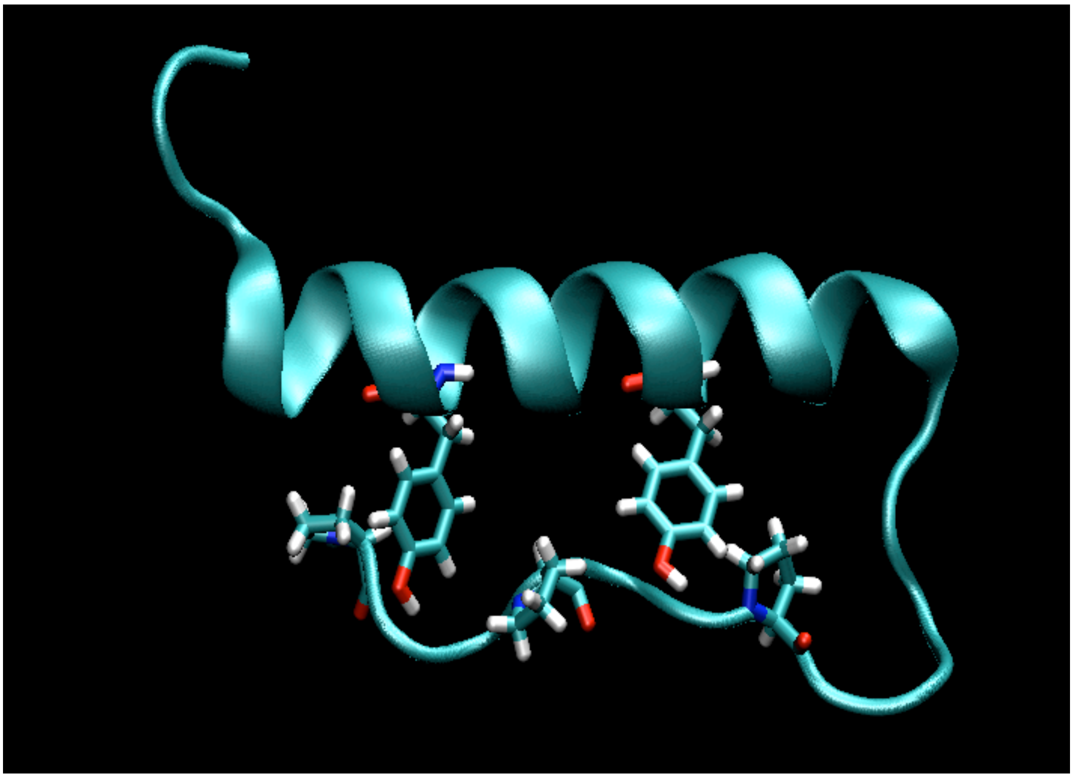


Figure 10: Refined structure of porcine PYY with an all atom representation of Pro and Tyr residues.

The last study was focused on folding of pPYY. By studying the formation of tertiary structure it was obvious that the peptide acts as a two state folder, and that breaking of several important contacts leads to disruption of the tertiary structure. The chemical shift changes can serve as sensitive probes to measure the thermal (un)folding process. The majority of chemical shifts can be traced throughout the folding process. The folded and unfolded states were in fast equilibrium, so that the observed chemical shifts correspond to a population-weighted average over these two states. The possibility to observe the concerted changes on many sites together, with their accompanying thermodynamical properties, delivered highly valuable and detailed data on the folding process. In particular, we have measured denaturation curves for each of the observed $^{13}C\alpha$ chemical shifts and calculated the inflection points, from which we could derive the temperature of unfolding for all sites. The data confirm, that PYY folding/unfolding is highly cooperative. In addition, the methanol denaturation data for pPYY provide detailed information on the effects of this solvent on folding. The solvent was gradually changed from pure water to pure methanol, and the $^{13}C\alpha$ chemical shifts were monitored during this process. It should be noted that unfolding in methanol was principally different in nature when compared to thermal unfolding in water: Unfolding in water leads to the complete loss of structure whereas unfolding in methanol results in removal of tertiary structure while secondary structure persists or is even stabilized. The comparison of the structure of PYY in methanol to that in aqueous solutions of DPC micelles reveals that their secondary structure were closely similar to each other. In both environments no tertiary structure is observed, and the C-terminal α-helix extends over the same region. Methanol therefore seems to be a good mimic of the environment of a water-membrane interfacial region, and we have proposed the usage of water-methanol mixtures to mimic the transition when the peptide traverses from bulk solution into the interface.

References

1. Sela, M., White, F. H. & Anfinsen, C. B. (1957) Reductive cleavage of disulfide bridges in ribonuclease, *Science. 125*, 691-2.
2. Dodson, E. J. (2007) Comput. biol.: Protein predictions, *Nature. 450*, 176-177.
3. Korzhnev, D. M., Salvatella, X., Vendruscolo, M., Nardo, A. A. D., Davidson, A. R., Dobson, C. M. & Kay, L. E. (2004) Low-populated folding intermediates of Fyn SH3 characterized by relaxation dispersion NMR, *Nature. 430*, 586-590.
4. Dobson, C. M. (2004) Principles of protein folding, misfolding and aggregation, *Semin Cell Dev Biol. 15*, 3-16.
5. Anfinsen, C. B. (1973) Principles that govern the folding of protein chains, *Science. 181*, 223-30.
6. Privalov, P. L., Gill, S. J. & C.B. Anfinsen, J. T. E. (1988) Stability of Protein Structure and Hydrophobic Interaction in pp. 191-234.
7. Linderstrøm-Lang, K. U. (1924) On the ionisation of proteins, *C.R. Trav. Lab. Carlsberg. 15*, 1-29.
8. Mirsky, A. E. & Pauling, L. (1936) On the Structure of Native, Denatured, and Coagulated Proteins, *Proc Natl Acad Sci USA. 22*, 439-47.
9. Eyring, H. & Stearn, A. E. (1939) The Application of the Theory of Absolute Reacton Rates to Proteins, *Chem Rev. 24*, 253-270.
10. Honig, B. H., Hubbell, W. L. & Flewelling, R. F. (1986) Electrostatic Interactions in Membranes and Proteins *Annu Rev Biophys Bio. 15*, 163-193.
11. Honig, B. H. & Hubbell, W. L. (1984) Stability of "salt bridges" in membrane proteins, *Proc Natl Acad Sci USA. 81*, 5412-6.
12. Pauling, L. (1960) *The Nature of the Chemical Bond and the Structure of Molecules and Crystals; An Introduction to Modern Structural Chemistry*, Cornell University Press.
13. Kauzmann, W. (1959) Some factors in the interpretation of protein denaturation, *Adv Protein Chem. 14*, 1-63.
14. Baker, E. N. & Hubbard, R. E. (1984) Hydrogen bonding in globular proteins, *Prog Biophys Mol Bio. 44*, 97-179.
15. Linderstrøm-Lang, K. U. (1952) Proteins and enzymes in *Lane Medical Lectures*
16. Kauzmann, W. (1954) The mechamism of enzyme action in *A Symposium of the the mechamism of enzyme action*, John Hopkins University Press, Baltimore.

17. Singer, S. J. & C.B. Anfinsen, Jr. (1963) The Properties of Proteins in Nonaqueous Solvents in pp. 1-68 Academic Press.
18. Hilderbrand, J. H. & S., R. L. (1964) *The solubility of nonelectrolytes*, Dover publications.
19. Fersht, A. R. (1995) Optimization of rates of protein folding: The nucleation-condensation mechanism and its implications, *Proc Natl Acad Sci USA. 92*, 10869-10873.
20. Ptitsyn, O. B. (1973) [Stages in the mechanism of self-organization of protein molecules], *Dokl Akad Nauk+. 210*, 1213-5.
21. Kim, P. S. & Baldwin, R. L. (1982) Specific intermediates in the folding reactions of small proteins and the mechanism of protein folding, *Annu Rev Biochem. 51*, 459-489.
22. Robson, B. & Pain, R. H. (1971) Analysis of the code relating sequence to conformation in proteins: Possible implications for the mechanism of formation of helical regions, *J Mol Biol. 58*, 237-257.
23. Crick, F. H. C. (1953) The packing of helices: simple coiled-coils, *Acta Crystallogr. 6*, 689-697.
24. Bryngelson, J. D., Wolynes, P.G. (1987) Spin glasses and the statistical mechanics of protein folding, *Proc Natl Acad Sci USA. 84*, 7524-7528.
25. Schultz, C. P. (2000) Illuminating folding intermediates, *Nat Struct Mol Biol. 7*, 7-10.
26. Dobson, C. M. (2003) Protein folding and misfolding, *Nature. 426*, 884-90.
27. Sugase, K., Dyson, H. J. & Wright, P. E. (2007) Mechanism of coupled folding and binding of an intrinsically disordered protein, *Nature. 447*, 1021-1025.
28. Dyson, H. J. & Wright, P. E. (2005) Intrinsically unstructured proteins and their functions, *Nat Rev Mol Cell Biol. 6*, 197-208.
29. Koshland, D. E. (1958) Application of a Theory of Enzyme Specificity to Protein Synthesis, *Proc Natl Acad Sci USA. 44*, 98-104.
30. Lange, O. F., Lakomek, N., Fares, C., Schroder, G. F., Walter, K. F. A., Becker, S., Meiler, J., Grubmuller, H., Griesinger, C. & Groot, B. L. d. (2008) Recognition Dynamics Up to Microseconds Revealed from an RDC-Derived Ubiquitin Ensemble in Solution, *Science. 320*, 1471-1475.
31. Iakoucheva, L. M., Radivojac, P., Brown, C. J., O'Connor, T. R., Sikes, J. G., Obradovic, Z. & Dunker, A. K. (2004) The importance of intrinsic disorder for protein phosphorylation, *Nucl. Acids Res. 32*, 1037-1049.
32. Pawar, A. P., DuBay, K. F., Zurdo, J., Chiti, F., Vendruscolo, M. & Dobson, C. M. (2005) Prediction of "Aggregation-prone" and "Aggregation-susceptible" Regions in Proteins Associated with Neurodegenerative Diseases, *J Mol Biol. 350*, 379-392.
33. Venkatraman, P., Wetzel, R., Tanaka, M., Nukina, N. & Goldberg, A. L. (2004) Eukaryotic Proteasomes Cannot Digest Polyglutamine Sequences and Release Them during Degradation of Polyglutamine-Containing Proteins, *Mol Cell. 14*, 95-104.
34. Dobson, C. M. (2002) Protein-misfolding diseases: Getting out of shape, *Nature. 418*, 729-730.
35. Buchner, J. (2002) Introduction: the cellular protein folding machinery, *Cell Mol Life Sci. 59*, 1587-1588.
36. Pickart, C. M. & Eddins, M. J. (2004) Ubiquitin: structures, functions, mechanisms, *Biochim Biophys Acta. 1695*, 55-72.
37. Bayrhuber, M., Ryazanov, S., Meins, T., Villinger, S., Habeck, M., Vonrhein, C., Walter, K., Lee, D., Karpinar, D., Grimm, S., Becker, S., Lange, A., Zweckstetter, M., Leonov, A., Zeth, K. & Griesinger, C. (2008). From Parkinson's disease to membrane

proteins. Presented at the *Conference*.

38. Weiss, S. (2000) Measuring conformational dynamics of biomolecules by single molecule fluorescence spectroscopy, *Nat Struct Mol Biol. 7*, 724-729.

39. Valeur, B. (2001) *Molecular Fluorescence: Principles and Applications*.

40. Förster, T. (1948) Zwischenmolekulare Energiewanderung und Fluoreszenz, *Ann Phys. 437*, 55-75.

41. Yang, H., Luo, G., Karnchanaphanurach, P., Louie, T., Rech, I., Cova, S., Xun, L. & Xie, X. S. (2003) Protein Conformational Dynamics Probed by Single-Molecule Electron Transfer, *Science. 302*, 262-266.

42. Wagman, M. E., Dobson, C. M. & Karplus, M. (1980) Proton NMR studies of the association and folding of glucagon in solution, *FEBS Lett. 119*, 265-270.

43. Campbell, I. D., Dobson, C. M., Moore, G. R., Perkins, S. J. & Williams, R. J. P. (1976) Temperature dependent molecular motion of a tyrosine residue of ferrocytochrome C, *FEBS Lett. 70*, 96-100.

44. Roder, H., Elove, G. A. & Englander, S. W. (1988) Structural characterization of folding intermediates in cytochrome c by H-exchange labeling and proton NMR, *Nature. 335*, 700-704.

45. Udgaonkar, J. B. & Baldwin, R. L. (1988) NMR evidence for an early framework intermediate on the folding pathway of ribonuclease A, *Nature. 335*, 694-699.

46. Nishimura, C., Dyson, H. J. & Wright, P. E. (2005) Enhanced picture of protein-folding intermediates using organic solvents in H/D exchange and quench-flow experiments, *Proc Natl Acad Sci USA. 102*, 4765-70.

47. Eliezer, D., Chung, J., Dyson, H. & Wright, P. (2000) Native and Non-native Secondary Structure and Dynamics in the pH 4 Intermediate of Apomyoglobin, *Biochemistry. 39*, 2894-2901.

48. Schwarzinger, S., Mohana-Borges, R., Kroon, G. J., Dyson, H. J. & Wright, P. E. (2008) Structural characterization of partially folded intermediates of apomyoglobin H64F, *Protein Sci. 17*, 313-321.

49. Dyson, H. & Wright, P. (2004) Unfolded Proteins and Protein Folding Studied by NMR, *Chem Rev. 104*, 3607-3622.

50. Yao, J., Chung, J., Eliezer, D., Wright, P. & Dyson, H. (2001) NMR Structural and Dynamic Characterization of the Acid-Unfolded State of Apomyoglobin Provides Insights into the Early Events in Protein Folding, *Biochemistry. 40*, 3561-3571.

51. Palmer, A. G. (2001) NMR probes of molecular dynamics: overview and comparison with other techniques, *Annu Rev Biophys Biomol Struct. 30*, 129-55.

52. Kay, L. E., Torchia, D. A. & Bax, A. (1989) Backbone dynamics of proteins as studied by ^{15}N inverse detected heteronuclear NMR spectroscopy: application to staphylococcal nuclease, *Biochemistry. 28*, 8972-9.

53. Křížová, H., Žídek, L., Stone, M. J., Novotny, M. V. & Sklenář, V. (2004) Temperature-dependent spectral density analysis applied to monitoring backbone dynamics of major urinary protein-I complexed with the pheromone 2-sec-butyl-4,5-dihydrothiazole*, *J Biomol NMR. 28*, 369-384.

54. Lipari, G. & Szabo, A. (1982) Model-free approach to the interpretation of nuclear magnetic resonance relaxation in macromolecules. 1. Theory and range of validity, *J Am Chem Soc. 104*, 4546-4559.

55. Peng, J. W. & Wagner, G. (1992) Mapping of spectral density functions using heteronuclear NMR relaxation measurements, *J Mag Res (1969). 98*, 308-332.

56. Peng, J. W. & Wagner, G. (1995) Frequency Spectrum of NH Bonds in Eglin c from Spectral Density Mapping at Multiple Fields, *Biochemistry. 34*, 16733-16752.

57. Gardner, K. H. & Kay, L. E. (1998) The use of ^{2}H, ^{13}C, ^{15}N multidimensional

NMR to study the structure and dynamics of proteins, *Annu. Rev. Biophys. Biomol. Struct. 27*, 357-406.

58. Tugarinov, V. & Kay, L. E. (2004) An Isotope Labeling Strategy for Methyl TROSY Spectroscopy, *J Biomol NMR. 28*, 165-172.

59. Millet, O., Loria, J., Kroenke, C., Pons, M. & Palmer, A. G. (2000) The Static Magnetic Field Dependence of Chemical Exchange Linebroadening Defines the NMR Chemical Shift Time Scale, *J Am Chem Soc. 122*, 2867-2877.

60. Skrynnikov, N. R., Millet, O. & Kay, L. (2002) Deuterium Spin Probes of Side-Chain Dynamics in Proteins. 2. Spectral Density Mapping and Identification of Nanosecond Time-Scale Side-Chain Motions, *J Am Chem Soc. 124*, 6449-6460.

61. Bax, A. (2003) Weak alignment offers new NMR opportunities to study protein structure and dynamics, *Protein Sci. 12*, 1-16.

62. Bouvignies, G., Bernardo, P., Meier, S., Cho, K., Grzesiek, S., Bruschweiler, R. & Blackledge, M. (2005) Identification of slow correlated motions in proteins using residual dipolar and hydrogen-bond scalar couplings, *Proc Natl Acad Sci USA. 102*, 13885-13890.

63. Korzhnev, D. M. & Kay, L. E. (2008) Probing Invisible, Low-Populated States of Protein Molecules by Relaxation Dispersion NMR Spectroscopy: An Application to Protein Folding, *Acc Chem Res. 41*, 442-451.

64. Carr, H. Y. & Purcell, E. M. (1954) Effects of Diffusion on Free Precession in Nuclear Magnetic Resonance Experiments, *Phys Rev. 94*, 630.

65. Hahn, E. L. (1950) Spin Echoes, *Phys Rev. 80*, 580.

66. McConnell, H. M. (1958) Reaction Rates by Nuclear Magnetic Resonance, *J Chem phys. 28*, 430-431.

67. Abergel, D. & III, A. G. P. (2003) On the use of the stochastic Liouville equation in nuclear magnetic resonance: Application to relaxation in the presence of exchange, *Concept Magn Res A. 19A*, 134-148.

68. Lundström, P., Hansen, D. F. & Kay, L. (2008) Measurement of carbonyl chemical shifts of excited protein states by relaxation dispersion NMR spectroscopy: comparison between uniformly and selectively ^{13}C labeled samples, *J Biomol NMR. 42*, 35-47.

69. Tollinger, M., Skrynnikov, N. R., Mulder, F., Forman-Kay, J. & Kay, L. (2001) Slow Dynamics in Folded and Unfolded States of an SH3 Domain, *J Am Chem Soc. 123*, 11341-11352.

70. Mittermaier, A., Korzhnev, D. & Kay, L. (2005) Side-Chain Interactions in the Folding Pathway of a Fyn SH3 Domain Mutant Studied by Relaxation Dispersion NMR Spectroscopy, *Biochemistry. 44*, 15430-15436.

71. Fersht, A. (1998) *Structure and Mechanism in Protein Science: A Guide to Enzyme Catalysis and Protein Folding.*

72. Neudecker, P., Zarrine-Afsar, A., Choy, W., Muhandiram, D. R., Davidson, A. R. & Kay, L. E. (2006) Identification of a Collapsed Intermediate with Non-native Long-range Interactions on the Folding Pathway of a Pair of Fyn SH3 Domain Mutants by NMR Relaxation Dispersion Spectroscopy, *J Mol Biol. 363*, 958-976.

73. Neudecker, P., Zarrine-Afsar, A., Davidson, A. R. & Kay, L. E. (2007) Φ-value analysis of a three-state protein folding pathway by NMR relaxation dispersion spectroscopy, *Proc Natl Acad Sci USA. 104*, 15717-22.

74. Daniel, R. M., Dunn, R. V., Finney, J. L. & Smith, J. C. (2003) The role of dynamics in enzyme activity, *Annu Rev Biophys Biomol Struct. 32*, 69-92.

75. Hill, R., Bracken, C., DeGrado, W. & Palmer, A. G. (2000) Molecular Motions and Protein Folding: Characterization of the Backbone Dynamics and Folding

Equilibrium of α_2D Using ^{13}C NMR Spin Relaxation, *J Am Chem Soc. 122*, 11610-11619.

76. Mulder, F. A., Tilborg, P. J. v., Kaptein, R. & Boelens, R. (1999) Microsecond time scale dynamics in the RXR DNA-binding domain from a combination of spin-echo and off-resonance rotating frame relaxation measurements, *J Biomol NMR. 13*, 275-288.

77. Farrow, N. A., Zhang, O., Forman-Kay, J. D. & Kay, L. E. (1995) Comparison of the Backbone Dynamics of a Folded and an Unfolded SH3 Domain Existing in Equilibrium in Aqueous Buffer, *Biochemistry. 34*, 868-878.

78. Farrow, N. A., Zhang, O., Szabo, A., Torchia, D. A. & Kay, L. E. (1995) Spectral density function mapping using ^{15}N relaxation data exclusively, *J Biomol NMR. 6*, 153-162.

79. Kim, S., Bracken, C. & Baum, J. (1999) Characterization of millisecond time-scale dynamics in the molten globule state of α-lactalbumin by NMR, *J Mol Biol. 294*, 551-560.

80. Farrow, N. A., Zhang, O., Forman-Kay, J. D. & Kay, L. E. (1994) A heteronuclear correlation experiment for simultaneous determination of ^{15}N longitudinal decay and chemical exchange rates of systems in slow equilibrium, *J Biomol NMR. 4*, 727-734.

81. Tolman, J. R., Flanagan, J. M., Kennedy, M. A. & Prestegard, J. H. (1995) Nuclear magnetic dipole interactions in field-oriented proteins: information for structure determination in solution, *Proc Natl Acad Sci USA. 92*, 9279-83.

82. Tjandra, N. & Bax, A. (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium, *Science. 278*, 1111-4.

83. Shortle, D. & Ackerman, M. S. (2001) Persistence of Native-Like Topology in a Denatured Protein in 8 M Urea, *Science. 293*, 487-489.

84. Ackerman, M. & Shortle, D. (2002) Molecular Alignment of Denatured States of Staphylococcal Nuclease with Strained Polyacrylamide Gels and Surfactant Liquid Crystalline Phases, *Biochemistry. 41*, 3089-3095.

85. Vallurupalli, P., Hansen, D. F., Stollar, E., Meirovitch, E. & Kay, L. E. (2007) Measurement of bond vector orientations in invisible excited states of proteins, *Proc Natl Acad Sci USA. 104*, 18473–18477.

86. Loria, J. P., Rance, M. & Palmer, A. G. (1999) A TROSY CPMG sequence for characterizing chemical exchange in large proteins, *J Biomol NMR. 15*, 151-155.

87. Skrynnikov, N. R., Dahlquist, F. & Kay, L. (2002) Reconstructing NMR Spectra of "Invisible" Excited Protein States Using HSQC and HMQC Experiments, *J Am Chem Soc. 124*, 12352-12360.

88. Lietzow, M. A., Jamin, M., Dyson, H. J. & Wright, P. E. (2002) Mapping Long-range Contacts in a Highly Unfolded Protein, *J Mol Biol. 322*, 655-662.

89. Hubbell, W. L., Mchaourab, H. S., Altenbach, C. & Lietzow, M. A. (1996) Watching proteins move using site-directed spin labeling, *Structure. 4*, 779-83.

90. Schmidt, P. G. & Kuntz, I. D. (1984) Distance measurements in spin-labeled lysozyme, *Biochemistry. 23*, 4261-4266.

91. Felitsky, D. J., Lietzow, M. A., Dyson, H. J. & Wright, P. E. (2008) Modeling transient collapsed states of an unfolded protein to provide insights into early folding events, *Proc Natl Acad Sci USA. 105*, 6278-83.

92. Grzesiek, S., Allan, M., Cordier, F., Dames, S., Duma, L., Nisius, L., Sass, J., Strohmeier, M., Vajpai, N. & Zähringer, U. (2008). NMR as a quantitative tool to elucidate biomolecular structure and dynamics. Presented at the *Conference*.

93. Gunsteren, W. F. v., Bakowies, D., Baron, R., Chandrasekhar, I., Christen, M.,

Daura, X., Gee, P., Geerke, D. P., Glättli, A., Hönenberger, P. H., Kastenholz, M. A., Oostenbrink, C., Schenk, M., Trzesniak, D., Vegt, N. F. A. v. d. & Yu, H. B. (2006) Biomolecular Modeling: Goals, Problems, Perspectives, *Angew. Chem. Int. Ed. 45*, 4064-4092.

94. Gunsteren, W. F. v. (2007) Foldamers (First Edition) pp. 173-192.

95. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*.

96. Bax, A. (1989) Two-dimensional NMR and protein structure, *Annu Rev Biochem. 58*, 223-56.

97. Güntert, P., Mumenthaler, C. & Wüthrich, K. (1997) Torsion angle dynamics for NMR structure calculation with the new program D, *J Mol Biol. 273*, 283-298.

98. Gunsteren, W. F. v., Dolenc, J. & Mark, A. E. (2008) Molecular simulation as an aid to experimentalists, *Curr Opin Struct Biol. 18*, 149-153.

99. Keller, B., Christen, M., Oostenbrink, C. & Gunsteren, W. F. v. (2007) On using oscillating time-dependent restraints in MD simulation, *J Biomol NMR. 37*, 1-14.

100. Daura, X., Antes, I., Gunsteren, W. F. v., Thiel, W. & Mark, A. E. (1999) The effect of motional averaging on the calculation of NMR-derived structural properties, *Proteins. 36*, 542-555.

101. Scott, W. R., Mark, A. E. & Gunsteren, W. F. v. (1998) On using time-averaging restraints in molecular dynamics simulation, *J Biomol NMR. 12*, 501-508.

102. Zagrovic, B. & Gunsteren, W. F. v. (2006) Comparing atomistic simulation data with the NMR experiment: how much can NOEs actually tell us?, *Proteins. 63*, 210-8.

103. Glattli, A., Daura, X., Seebach, D. & Gunsteren, W. F. v. (2002) Can One Derive the Conformational Preference of a β-Peptide from Its CD Spectrum?, *J Am Chem Soc. 124*, 12972-12978.

104. Hummer, G. & Szabo, A. (2005) Free Energy Surfaces from Single-Molecule Force Spectroscopy, *Acc Chem Res. 38*, 504-513.

105. Daura, X., Gunsteren, W. F. v., Rigo, D., Jaun, B. & Seebach, D. (1997) Studying the Stability of a Helical β-Heptapeptide by Molecular Dynamics Simulations, *Chem-Eur J. 3*, 1410-1417.

106. Daura, X., Jaun, B., Seebach, D., Gunsteren, W. F. v. & Mark, A. E. (1998) Reversible peptide folding in solution by molecular dynamics simulation, *J Mol Biol. 280*, 925-932.

107. Christen, M., Keller, B. & Gunsteren, W. F. v. (2007) Biomolecular structure refinement based on adaptive restraints using local-elevation simulation, *J Biomol NMR. 39*, 265-273.

108. Huber, T., Torda, A. E. & Gunsteren, W. F. v. (1994) Local elevation: A method for improving the searching properties of molecular dynamics simulation, *J Comput Aided Mol Des. 8*, 695-708.

109. Kony, D. B., Hunenberger, P. H. & Gunsteren, W. F. v. (2007) Molecular dynamics simulations of the native and partially folded states of ubiquitin: Influence of methanol cosolvent, pH, and temperature on the protein structure and dynamics, *Protein Sci. 16*, 1101-1118.

110. Boned, R., Gunsteren, W. F. v. & Daura, X. (2008) Estimating the Temperature Dependence of Peptide Folding Entropies and Free Enthalpies from Total Energies in Molecular Dynamics Simulations, *Chem-Eur J. 14*, 5039-5046.

111. Christen, M. & Gunsteren, W. F. v. (2008) On searching in, sampling of, and dynamically moving through conformational space of biomolecular systems: A review, *J Comput Chem. 29*, 157-166.

112. Brünger, A. T., Krukowski, A. & Erickson, J. W. (1990) Slow-cooling protocols for crystallographic refinement by simulated annealing, *Acta Crystallogr A. 46*, 585-

93.

113. Ewald, P. P. (1921) Die Berechnung optischer und elektrostatischer Gitterpotentiale, *Ann Phys. 369*, 253-287.

114. Ferrara, P. & Caflisch, A. (2000) Folding simulations of a three-stranded antiparallel β-sheet peptide, *Proc Natl Acad Sci USA. 97*, 10780-5.

115. Daura, X., Mark, A. E. & Gunsteren, W. F. v. (1998) Parametrization of aliphatic CH_n united atoms of GROMOS96 force field, *J Comput Chem. 19*, 535-547.

116. Levitt, M. (1983) Protein folding by restrained energy minimization and molecular dynamics, *Journal of Molecular Biology. 170*, 723-64.

117. Torrie, G. M. & Valleau, J. P. (1977) Nonphysical sampling distributions in Monte Carlo free-energy estimation: Umbrella sampling, *J Comput Phys. 23*, 187-199.

118. Schlitter, J., Engels, M. & Krüger, P. (1994) Targeted molecular dynamics: a new approach for searching pathways of conformational transitions, *J Mol Graphics. 12*, 84-9.

119. Erp, T. S. v., Moroni, D. & Bolhuis, P. G. (2003) A novel path sampling method for the calculation of rate constants, *The J Chem Phys. 118*, 7762-7774.

120. Bolhuis, P. G., Chandler, D., Dellago, C. & Geissler, P. L. (2002) Transition path sampling: throwing ropes over rough mountain passes, in the dark, *Ann Rev Phys Chem. 53*, 291-318.

121. Prigogine, I. (2002) Advances in Chemical Physics: 123 in *Book* pp. 1-78.

122. Mark, A. E. & Gunsteren, W. F. v. (1992) Simulation of the thermal denaturation of hen egg white lysozyme: trapping the molten globule state, *Biochemistry. 31*, 7745-7748.

123. Dellago, C., Bolhuis, P. G. & Chandler, D. (1998) Efficient transition path sampling: Application to Lennard-Jones cluster rearrangements, *J Chem Phys. 108*, 9236-9245.

124. Andersen, H. C. (1980) Molecular dynamics simulations at constant pressure and/or temperature, *J Chem Phys. 72*, 2384-2393.

125. Bolhuis, P. G. (2003) Transition path sampling on diffusive barriers, *J Phys-Condens Mat. 15*, S113-S120.

126. Bolhuis, P. G. (2003) Transition-path sampling of β-hairpin folding, *Proc Natl Acad Sci USA. 100*, 12129-34.

127. Moroni, D., Erp, T. S. v. & Bolhuis, P. G. (2004) Investigating rare events by transition interface sampling, *Physica A. 340*, 395-401.

128. Crehuet, R. & Field, M. J. (2003) A temperature-dependent nudged-elastic-band algorithm, *J Chem Phys. 118*, 9563-9571.

129. Bradley, P., Misura, K. M. S. & Baker, D. (2005) Toward High-Resolution de Novo Structure Prediction for Small Proteins, *Science. 309*, 1868-1871.

130. Simons, K. T., Bonneau, R., Ruczinski, I. & Baker, D. (1999) Ab initio protein structure prediction of CASP III targets using ROSETTA, *Proteins. 37*, 171-176.

131. Qian, B., Raman, S., Das, R., Bradley, P., McCoy, A. J., Read, R. J. & Baker, D. (2007) High-resolution structure prediction and the crystallographic phase problem, *Nature. 450*, 259-264.

132. Canutescu, A. A. & Dunbrack, R. L. (2003) Cyclic coordinate descent: A robotics algorithm for protein loop closure, *Protein Sci. 12*, 963-972.

133. Dunbrack-JR, R. L. & Cohen, F. E. (1997) Bayesian statistical analysis of protein side-chain rotamer preferences, *Protein Sci. 6*, 1661-1681.

134. Li, Z. & Scheraga, H. A. (1987) Monte Carlo-minimization approach to the multiple-minima problem in protein folding, *Proc Natl Acad Sci USA. 84*, 6611-5.

135. Popot, J. L. & Engelman, D. M. (1990) Membrane protein folding and oligomerization: the two-stage model, *Biochemistry. 29*, 4031-7.
136. Bowie, J. U. (2005) Solving the membrane protein folding problem, *Nature. 438*, 581-589.
137. Barth, P., Schonbrun, J. & Baker, D. (2007) Toward high-resolution prediction and design of transmembrane helical protein structures, *Proc Natl Acad Sci USA. 104*, 15682–15687.
138. Jiang, L., Althoff, E. A., Clemente, F. R., Doyle, L., Rothlisberger, D., Zanghellini, A., Gallaher, J. L., Betker, J. L., Tanaka, F., Barbas, C. F., Hilvert, D., Houk, K. N., Stoddard, B. L. & Baker, D. (2008) De Novo Computational Design of Retro-Aldol Enzymes, *Science. 319*, 1387-1391.
139. Ashworth, J., Havranek, J. J., Duarte, C. M., Sussman, D., Monnat, R. J., Stoddard, B. L. & Baker, D. (2006) Computational redesign of endonuclease DNA binding and cleavage specificity, *Nature. 441*, 656-659.
140. Lerch, M., Kamimori, H., Folkers, G., Aguilar, M., Beck-Sickinger, A. & Zerbe, O. (2005) Strongly Altered Receptor Binding Properties in PP and NPY Chimeras Are Accompanied by Changes in Structure and Membrane Binding, *Biochemistry. 44*, 9255-9264.
141. Karplus, M. (1959) Contact Electron-Spin Coupling of Nuclear Magnetic Moments, *J. Phys. Chem. 33*, 11-15.

5. Probing the formation of stable tertiary structure in a model mini-protein at atomic resolution: Determinants of stability of a helical hairpin

The minimal model system to study the basic principles of protein folding is the hairpin. The formation of β-hairpins, which are the basic components of antiparallel β-sheets, has been studied extensively in the past decade but much less is known about helical hairpins. Here, we probe hairpin formation between a polyproline type-II helix and an α-helix as present in the natural mini-protein peptide YY (PYY). Both turn sequence and interactions of aromatic side-chains from the C-terminal α-helix with the pockets formed by N-terminal Pro residues are shown by site-directed mutagenesis and solution NMR spectroscopy in different solvent systems to be important determinants of backbone dynamics and hairpin stability, suggesting a close analogy with some β-hairpin structures. It is shown that multiple relatively weak contacts between the helices are necessary for the formation of the helical hairpin studied here, whilst the type I β-turn acts like a hinge, which through certain single amino acid substitutions is destabilized such that hairpin formation is completely abolished. Denaturation and renaturation of tertiary structure by temperature or co-solvents were probed by measuring changes of chemical shifts. Folding of PYY is both reversible and cooperative as inferred from the sigmoidal denaturation curves displayed by residues at the interface of the helical hairpin. Such mini-proteins thus feature an important hallmark of globular proteins and should provide a convenient system to study basic aspects of helical hairpin folding that are complementary to those derived from studies of β-hairpins.

published as: Neumoin, A., Mares, J., Lerch-Bader, M., Bader, R., Zerbe, O. (2007) *J. Am. Chem. Soc.* 129(28): 8811-7

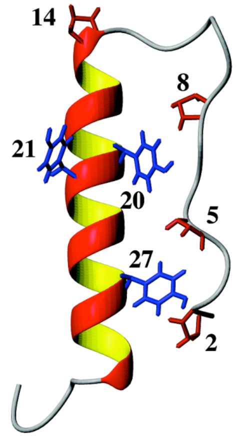
Introduction

Despite extensive efforts in studying folding of peptides and proteins, a much better understanding of molecular determinants for particular folds and their formation is still highly desirable. While polypeptide chains that fold into specific secondary structures are comparatively easy to design, the construction of proteins with defined tertiary structure, good side-chain packing and cooperative folding behavior is still a major challenge. The reason for this difficulty is that tertiary contacts are often stabilized by a manifold of weaker interactions clearly underlining the necessity for a deeper understanding of how such contacts are made and stabilized.

The smallest natural domains that fold autonomously into protein-like structures have 32-40 residues [1, 2], although stable β-hairpin structures with a much smaller number of residues have successfully been isolated from a larger protein [3] or even designed de novo [4]. The availability of small and well-folded β-hairpin structures is presently advancing our understanding of factors that govern protein folding at ever-increasing pace (see e.g. Hughes et al. [5] for a review). Particular interest has been paid to β-hairpins that display cooperative folding behavior, since this is a hallmark of natural proteins [6]. It has been found that the stability of β-hairpins is determined by contributions from turn [7, 8], intrinsic sheet propensities [9] and hydrophobic side-chain interactions across the strands [4]. In fact, the relative contributions of intrinsic conformational bias and interstrand side-chain – side-chain effects are suggested to be of comparable magnitude [9].

While the formation of individual α-helices and β-hairpins has been studied in great detail, relatively little is known about the factors contributing to helical hairpin formation. In the absence of additional covalent constraints (such as disulfide bonds [10, 11]) helical hairpins have been reported to be marginally stable [12] or to oligomerize into larger assemblies such as helical bundles [13]. One of the few sequences shown to fold into a monomeric and stable hairpin in solution is α t α [14]. Although the hydrophobic interface appears to contribute significantly to the stability of the two helices in α t α, thermal unfolding was surprisingly uncooperative. This is indicative of a lack of specific tertiary interactions (by either hydrogen-bonding or interdigitating side-chains) and suggests that α t α represents a molten globule intermediate rather than a protein-like structure.

Here we present a detailed investigation on structural determinants of cooperative helical hairpin formation in 36 residue peptides, which are derived from the neuropeptide Y family of peptide hormones. Some of these peptides adopt a well-defined hairpin structure in water, which was first observed by Blundell [15] for avian pancreatic peptide (aPP) using X-ray crystallography. This surprisingly stable helical hairpin is commonly referred to as PP-fold and is characterized by a C-terminal α-helix, which is back-folded via a type I β-turn onto an N-terminal polyproline type-II helix. Tertiary contacts between the α-helix and the polyproline helix are shown in Fig.1:



1 11 21 31

pPYY YPAKP EAPGE DASPE ELSRY YASLR HYLNL VTRQR Y NH_2

pNPY YPSKP DNPGE DAPAE DLARY YSALR HYINL ITRQR Y NH_2

bPP APLEP EYPGD NATPE QMAQY AAELR RYINM LTRPR Y NH_2

Figure 1: Structure of porcine PYY with side-chains of Pro and Tyr residues depicted in red or blue, respectively. The sequences of porcine PYY and NPY and bovine PP are displayed on the right. Residues that are conserved among different species are underlined, and residues, which are part of the present mutational analysis, are depicted in red.

The solution structure of the peptide YY (PYY) was shown to possess a highly similar helical hairpin [16, 17] whilst tertiary structure is surprisingly absent in the highly homologous neuropeptide Y (NPY) [18]. Despite its small size the peptide YY is stably folded in aqueous solutions, as evidenced by non-averaged NMR parameters, a high-resolution NMR structure and uniformly low internal backbone dynamics, as independently verified from NMR relaxation experiments [17]. To probe the energetic significance of the tertiary contacts we replaced residues at the hydrophobic interface of the hairpin-type structure of PYY by Ala and analyzed by high-resolution NMR spectroscopy the conformational and dynamical properties of the resulting peptides. By varying temperature or the methanol content of the aqueous solvent and monitoring chemical shifts we followed the residue-specific formation of tertiary contacts while changing the physical or chemical environment. The methods used by us efficiently deliver structural and dynamical information on changes in the stabilities and folding behavior of the different mutants at atomic resolution and hence provide a much more precise picture as compared to optical methods such as CD or fluorescence spectroscopy, which are usually limited to monitoring changes on molecular scales only.

The results suggest that helical hairpin formation in PP-fold peptides is both reversible and cooperative and that specific N- and C-terminal tertiary hydrophobic contacts between the polyproline and the α-helix provide the driving force for folding. In addition, structural analysis of substitutions in the turn region indicate that the loop behaves like a hinge, which may (or may not) favor, but does not constrain the hairpin structure. To our knowledge, the present data provide the first detailed investigation on cooperative tertiary structure formation in a natural, stably folded, but otherwise unconstrained helical hairpin and show that PP-fold mini-proteins can serve as a convenient system to study aspects of folding that are complementary to those using β-hairpins.

5.1 Results

The role of residues at the PP-fold interface

In our attempt to understand the contributions to stability of the PP-fold, we first chose to examine contacts at the back-fold interface made between Pro and aromatic Tyr residues. Recently, π-Pro interactions were postulated to contribute substantially to the stability of long-range contacts [19, 20]. Thus we replaced Pro residues 2,5 and 8 as well as Tyr20, 21 and 27 by Ala to produce all single mutations as well as the quadruple Ala2,5,8,14-PYY mutant by site-directed mutagenesis. The destabilization of the helical hairpin in these mutants is probed by measuring changes in internal backbone dynamics by means of the heteronuclear NOE between the amide nitrogen and its directly attached proton. Fig. 2A displays values of the $^{15}N\{^{1}H\}$-NOE for PYY in solution and when bound to DPC micelles. In the latter environment formation of tertiary structure is efficiently suppressed, demonstrating its utility for quantifying the stability of tertiary structure in the system under study [21, 22].

The $^{15}N\{^{1}H\}$-NOE data indicate that none of the Pro mutants are stably back-folded (see Fig. 2B). For the quadruple Pro mutant and NPY the $^{3}J_{HN\alpha}$ scalar couplings are > 6 Hz for all residues in the N-terminal segment, which is indicative of conformational averaging. By contrast, in the single Pro mutants the N-terminal segment is not fully flexible (see Supp. Mat.). Interestingly, the rigidity of the C-terminal α-helix is highly correlated to the rigidity of the turn and the adjacent amino acids encompassing residues 6 to 12. Moreover, the lack of long-range contacts destabilizes in particular the last two turns of the helix.

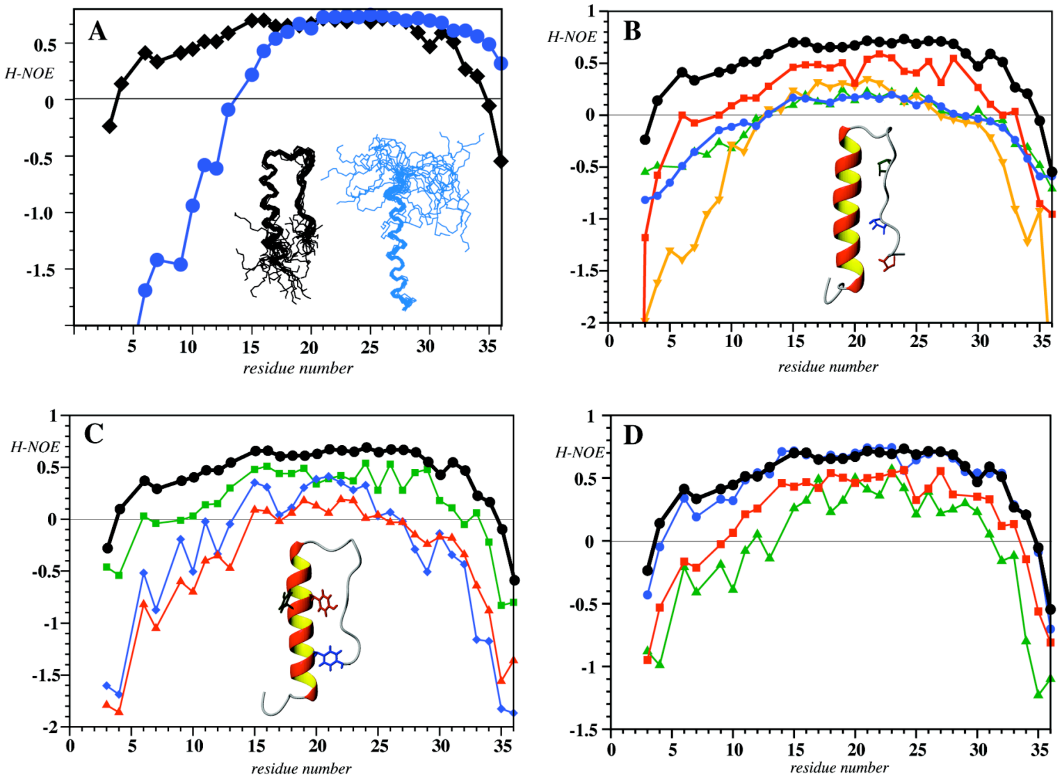


Figure 2: Values of the $^{15}N\{^1H\}$-NOE. All data were determined at 500 MHz in water using 1mM solutions of the peptides at pH 4.1 in 20mM acetate buffer. a) Effect of solvent system: Back-folded PYY in aqueous solution (black diamonds) and PYY when bound to DPC micelles (blue circles), where formation of the helical hairpin is suppressed. The corresponding structures displaying backbone atoms are depicted in the inset. b) Effect of proline mutations: PYY (black circles), Ala2-PYY(red squares), Ala5-PYY(blue circles), Ala8-PYY(green triangles) and Ala2,5,8,14-PYY(orange triangles). c) Effect of tyrosine mutations: PYY (black circles), Ala20-PYY (red triangles), Ala21-PYY(green squares) and Ala27-PYY (blue diamonds). d) Effect of mutations in the turn region: PYY(black circles), Ala14-PYY(blue circles), Pro13,Ala14-PYY(red squares) and Ala13-PYY (green triangles).

We have additionally investigated the structural role of Tyr20, 21 and 27 again by recording the $^{15}N\{^{1}H\}$-NOE values of the corresponding single Tyr->Ala mutants (Fig. 2C). The data confirm that these tyrosine residues and Pro2, 5 and 8 have complementary roles in the stabilization of both secondary and tertiary structure and support the data by Woll and Gellman on the Tyr20Ala and Tyr27Ala mutants [23]. In all the Tyr->Ala mutants, the back-fold is significantly destabilized. Mutation of Tyr20 or Tyr27, however, is far more destabilizing than mutation of Tyr21, because the π-systems of the former two residues point towards the back-folding interface and form contacts with Pro residues 5 and 8. Moreover, we have noticed in case of Ala27-PYY that the heteronuclear NOE drops in particular in the C-terminal part of the destabilized helix encompassing residues 24-30. In contrast, replacing Tyr20, which forms crucial contacts to both Pro5 and Pro8, results in a more general destabilization of the C-terminal helix. Considering that Tyr27 makes contacts with Pro2 removal of the former may result in fraying of the hairpin at its termini.

As the PYY mutants with a significantly destabilized PP-fold lack medium-range proton-proton NOEs in the N-terminal polyproline helix, we chose to probe for residual structure in the N-terminal segment by recording residual dipolar couplings in samples that were partly aligned in stretched polyacrylamide gels. The values measured for the weakly and strongly destabilized PYY mutants Ala2-PYY and Ala8-PYY, respectively, are depicted in Fig. 3 along with the RDCs from the stably back-folded wild-type PYY.

The comparison reveals that overall the RDCs of N-terminal residues of all three molecules follow a similar oscillatory pattern, although the amplitudes are reduced when the helical hairpin is destabilized (as inferred from the $^{15}N\{^{1}H\}$-NOE data). The magnitude of RDC values depends on the ensemble-averaged orientation of the NH bond vectors relative to the alignment tensor. Hence, the decreasing amplitudes indicate that the population of the helical hairpin is reduced in favor of species in which the N terminus is unstructured as seen both in Ala2-PYY and even more pronounced so in Ala8-PYY.

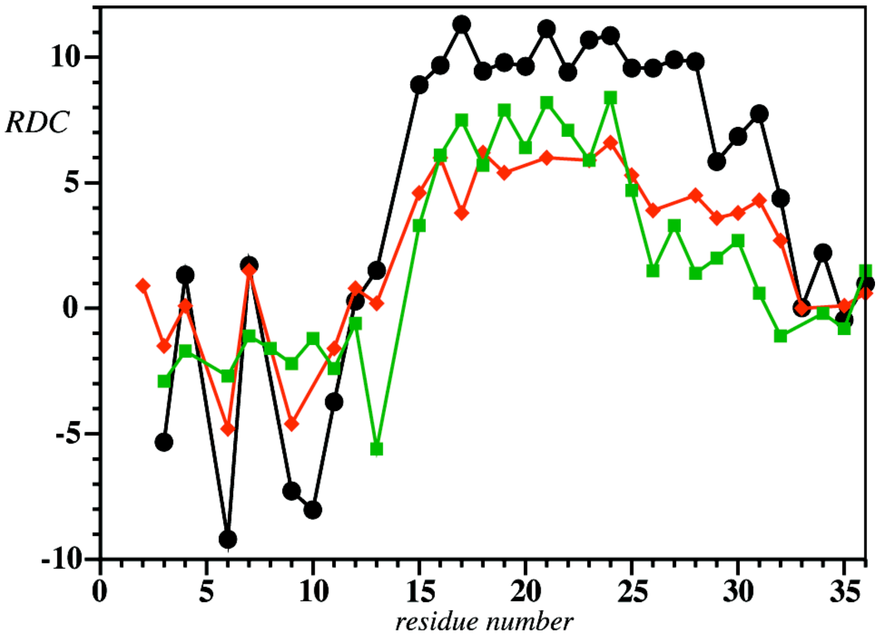


Figure 3: Residual dipolar couplings (RDCs) of PYY(black circles), Ala2-PYY(red diamonds) and Ala8-PYY(green squares) measured in stretched polyacrylamide gels at 700 MHz.

The differences in internal backbone dynamics of the N-terminal segment in different Pro->Ala mutants could possibly be related to i) differences in the propensities of the various mutants to adopt the PP-fold or ii) in changes in the intrinsic rigidity of this peptide segment due to the presence or absence of Pro residues. To resolve this ambiguity we have conducted measurements of the $^{15}N\{^{1}H\}$-NOE in DPC micelles, an environment, in which tertiary interactions are efficiently blocked. We have previously shown in case of bPP or pPYY by using micelle-integrating spin-labels that the amphipathic side of the helix becomes associated with the surface of the micelle, thereby releasing the N terminus [17, 24]. In the presence of micelles, the $^{15}N\{^{1}H\}$-NOE data of all single as well as the quadruple mutant are virtually identical (Supp. Mat Fig. S3). Moreover, the C-terminal helix of all peptides is stabilized on binding to micelles, whilst the H-NOE values of residues 6 to 14 indicate continuously increasing flexibility toward the N terminus. A comparison of the H-NOE of the quadruple mutant in water and in DPC micelles displays highly similar values for the N-terminal segment in both environments, providing evidence that the N terminus in this particular mutant behaves more or less like a free-flight chain also in solution. The data therefore suggest that the backbone rigidity of N-terminal residues of a given mutant in solution is mainly associated with its propensity to adopt the back-fold as a result of a number of specific tertiary contacts between aromatic residues of the C-terminal helix with proline residues of the N-terminal polyproline helix rather than structural properties of the N-terminal segment alone.

The importance of residues in the hinge region

The sequences of pNPY, pPYY and bPP are displayed in Figure 1 along with the structure of pPYY. Pro residues 2, 5 and 8 and Tyr residues 20 and 27, which together form the hydrophobic core of the folded hairpin, are conserved in all three peptides. It is therefore highly unlikely that these residues account for the different propensities of NPY and PYY to adopt the helical hairpin. The sequence homology of NPY and PYY is larger than 80%, and the most prominent difference is Pro found at position 14 in PYY while it occurs at position 13 in NPY. Proline 14 occupies a position in the turn region that links the C-terminal α-helix to the N-terminal polyproline helix. Due to the particular nature of the Pro side-chain we reasoned that Pro14 may direct the N-terminal segment into a favorable enough position for adopting the back-folded state. Accordingly, we have substituted Pro14 by Ala to relax this possible conformational restraint. A comparison of the structure of Ala14-pPYY with pPYY and bPP reveals that Ala14-pPYY very much resembles the structure of PYY or bPP in that the N-terminal segment is clearly back-folded (see Supp. Mat. Fig. S1) and similar interactions between C- and N-terminal residues occur. From these data we deduce that the PP-fold of Ala14-pPYY is sufficiently supported by the hydrophobic contacts made between Pro residues from the N terminus and Tyr residues from the C-terminal α-helix and that Pro14 is not needed to constrain the turn conformationally in order to enforce back-folding.

Although Pro14 apparently does not enforce tertiary contacts in PYY it may be that the otherwise highly homologous NPY cannot adopt the PP-fold because of a Pro in position 13, which is conserved in all NPY sequences known today. The data for the ^{15}N{^{1}H}-NOE of Pro13,Ala14-PYY are depicted in Fig. 2D and clearly prove that this mutant is not back-folded. A continuous decrease in magnitude of the ^{15}N{^{1}H}-NOE is observed for residues preceding Ala14. Negative H-NOE values are observed for N-terminal residues and the C-terminal helix is also significantly destabilized, but the values also indicate that the hairpin in this mutant is still more stable than in the quadruple mutant. By shifting Pro by one position towards the N terminus the naturally occurring Ser13-Pro14 is replaced by Pro13-Ala14. As a control for the effect of substituting Ser13 alone we also investigated hairpin formation by the single mutant Ala13-PYY. To our surprise we found that the PP-fold is even less stable in Ala13-PYY than in Pro13,Ala14-PYY (see Fig. 2D). An energy minimization of the turn region using an implicit solvent model (see Supp. Mat.) indicates the possible presence of hydrogen bonds of the hydroxyl group of Ser13 with the amide proton of Glu15 and to the side-chain carboxyl group of the latter residue. We conclude that the presence of a hydrogen bond donor at position 13 is crucial for correct formation of the β-turn, and that the shift of Pro from position 14 in PYY to position 13 in NPY or the replacement of Ser13 by Ala is sufficient to disrupt the tertiary contact. It is important to note that, although the amino acid composition of the turn region determines to what extent the turn may favor the formation of tertiary structure, it is nevertheless not sufficient to constrain the helical hairpin on its own. This is evidenced both by the fact that removal of a single tertiary contact in the hydrophobic cluster can fully disrupt the back-fold and by the backbone motional dynamics of the turn region, which is significantly more rigid when the hydrophobic cluster between N- and C-termini is stabilized (see above). On the other hand it is also clear that the turn rigidity weakens gradually under increasingly more destabilizing conditions. This fact is best seen in changes of the $^{3}J_{\mathrm{HN}\alpha}$ scalar couplings of residues from the turn region. For example, the coupling for residue Asp11 is always larger than 8 Hz (and often larger than 9 Hz) in the proline mutants in water but always close to 7 Hz in the DPC micelle-bound state, the latter value being indicative of conformational averaging. In addition, the value for this coupling for PYY in methanol is only moderately reduced to 8.7 Hz from the 10 Hz encountered in water. Taken together, these data indicate that the turn acts like a hinge that may support folding but does not provide sufficient stabilizing force to constrain the hairpin on its own.

Studying the formation of the helical hairpin

The stability of the folded state critically depends on the delicate balance between the free energies of the folded and unfolded states. Disrupting the helical hairpin in PYY will expose hydrophobic surfaces to the solvent. Accordingly, we have studied removal of tertiary structure in PYY both by thermal and by solvent-induced denaturation.

In the thermal denaturation experiments, proton spectra of 1mM solutions of PYY and Tyr7-PYY were measured in the temperature range from 280 to 370 K. The corresponding curves are characterized by sigmoidal shapes (see Fig. 4) and the melting points of residues from different regions of the polypeptide chain are within 1 degree for PYY and 1.2 degrees for Tyr7-PYY, suggesting a high degree of cooperativity. The data additionally reveal that the helical hairpin is slightly more stable in Tyr7-PYY than in PYY, as evidenced by an increase of the melting temperature from 316.5K in PYY to 326.5K in Tyr7-PYY:

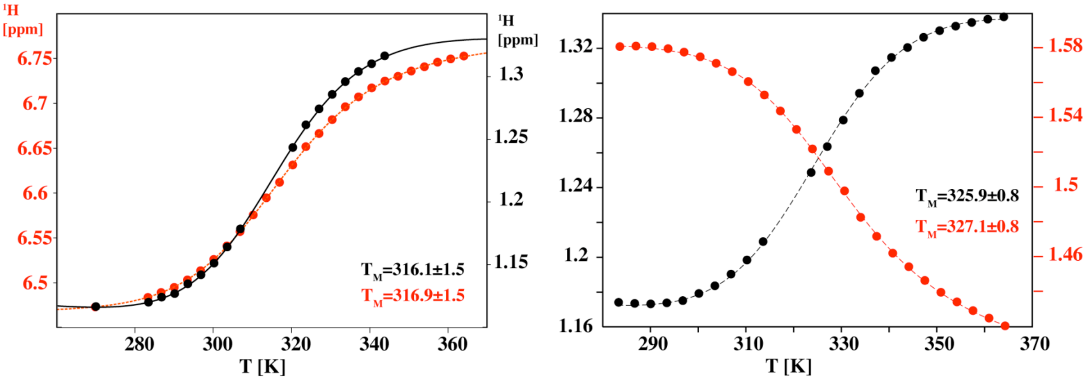


Figure 4: Left: Proton chemical shifts of the methyl group of Ala7 (black circles) and the Hδ of the aromatic system of Tyr20 (red circles) vs. temperature for PYY (left) and of methyl groups of Ala12 (black circles) and Leu17 (red circles) of Tyr7-PYY (right). The melting temperatures derived from fits to experimental data are denoted as T_M in the figure.

The increased PP-fold stability seen in Tyr7-PYY as compared to wildtype PYY must result from additional intra- or intermolecular interactions. It should be noted at this point that the related avian polypeptide (aPP) exists in dimeric form in the crystal structure [15] (K_d 0.32 μM at pH 5.0 [25]). Therein, the interface between the monomeric subunits is formed by a hydrophobic cluster, in which the π-systems of Tyr7, Tyr21 and Phe20 stack onto one another. By measuring chemical shift changes in a dilution series the K_d for homodimerization of PYY at pH 4.1 was determined to be 35±18 mM, indicating that the monomer concentration is between 91 and 95% at 1mM concentration. The chemical shift of the amide proton of Ala12 is very sensitive to the extent of hairpin formation, and the observed minor change of 0.06 ppm between 1mM and 10μM concentrations indicates that the dimer contributes only to a small extent to the stability of the helical hairpin. We therefore feel justified to treat the system as a monomer.

Apart from thermal denaturation, the PP-fold can also be destabilized by co-solvents with increased lipophilicity. It is known that co-addition of alcohols stabilizes helical secondary structures and disrupts weak hydrophobic contacts, which in turn unfolds tertiary structures. The order of effectiveness for destabilizing tertiary structure is trifluoroethanol > propanol > ethanol > methanol[26, 27]. The less dramatic changes observed with methanol indicate that this solvent may be particularly useful to monitor the back-folding transition[28]. Mixtures of water and methanol with methanol contents of about 30% have often been found to result in molten-globule-type structures that possess a considerable extent of native secondary structure while tertiary structure is largely destroyed[28-33].

To monitor the back-folding transition we have measured the amide proton chemical shifts as a function of the methanol content. The data for selected residues of Tyr7-PYY, for which the transition between non-back-folded and back-folded species occurs at a larger methanol-to-water ratio than for PYY (and hence is better visible) are displayed in Fig. 5.

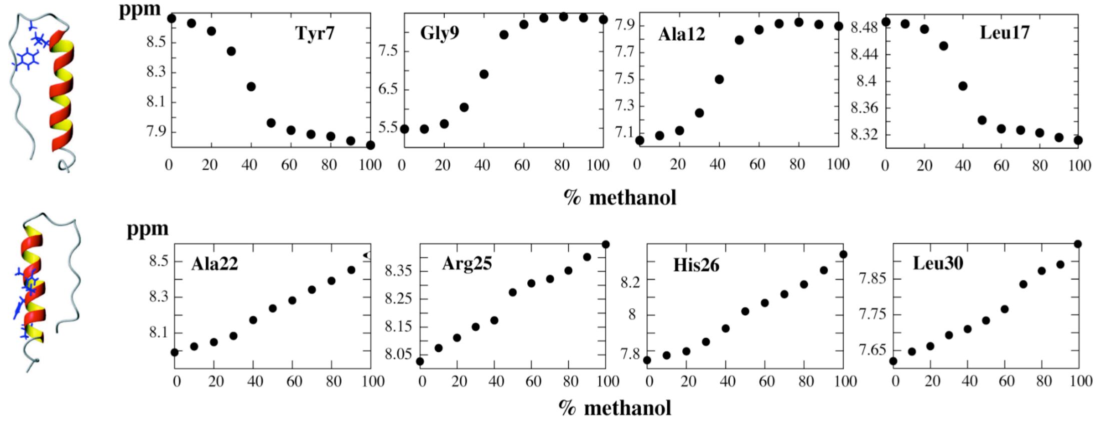


Figure. 5: Proton chemical shifts of selected amide protons of Tyr7-PYY *vs*. the methanol content for residues of the back-fold interface (top) and for those pointing away from it (bottom). On the left, side-chains of these residues have been drawn in the structures to indicate their location in the molecule.

The data clearly reveal that curves from residues of the back-fold interface display a characteristic sigmoidal shape. The point of inflection corresponds to a 41.8 ± 2.1% methanol/water mixture. In contrast, for residues of unstructured segments and for all residues that are pointing away from the back-fold interface the changes are limited to differences in solvent-coordination and/or stability of backbone hydrogen bonds and hence the curves appear approximately linear. The data clearly favor a model in which the PP-fold builds up *cooperatively*, because sigmoidal curves with similar inflection points are observed for the amide proton chemical shift changes for all the residues of the PP-fold interface. Similar sigmoidal curves are also seen for the values of the ${}^{3}J_{HN\alpha}$ coupling constants, which are directly dependent on the torsion angle ϕ (see Supp. Mat.). The associated free energy of formation of the helical hairpin in aqueous solution, computed following the procedure introduced by Santoro [34], is -17.4 ± 1.3 kJ/mol, which is in agreement with the values typically observed for the free energies of folding of globular proteins ranging from -20 to -60 kJ/mol [6] and similar to the value of -12.6 kJ/mol determined for bPP [23].

5.2 Discussion

In this study we have investigated molecular properties that lead to the cooperative formation of a helical hairpin in a natural mini-protein at atomic resolution. In particular, we have addressed the importance of the individual Pro and Tyr residues for forming tertiary structure in various mutants of PYY. Moreover, we characterized the significance of residues at specific positions in the turn region. The observation that each Pro->Ala replacement in the segment 2 to 8 reduces the stability of the back-fold significantly indicates that the simultaneous presence of all these contacts is required to form the hairpin. Measurements of backbone dynamics in the presence of DPC micelles, in which the formation of tertiary contacts between the helix and the N-terminal segment is efficiently prevented, revealed that differences in backbone dynamics between the various Pro->Ala mutants in the micelle-bound state are largely removed. Hence the increased rigidity is not primarily an intrinsic property of Pro-rich peptide segments but rather due to correct positioning of residues capable of stabilizing the back-folded conformation.

In a polyproline type II helix positions i and i+3 Pro residues point in the same direction. Moreover, the aromatic residues Tyr20 and Tyr27 are located at i,i+7 positions of a helix and therefore again point in the same direction. Pro residues and aromatic side-chains possess rather flat surfaces, enabling them to mutually intercalate in a zipper-type fashion resulting in a relatively large interaction surface. The importance of non-local Pro-aromatic interactions has been highlightened previously in different contexts, including interstrand stabilization of antiparallel β-sheets [21], a mini-protein called Trp cage [35, 36] and binding of proline-rich peptides to aromatic residues in SH3 domains [37]. It has indeed been speculated that Pro engages with an aromatic residue to form a C-H --- π interaction that may provide substantial binding energy [19, 20]. It is fully consistent with these observations, that the exchange of individual Tyr residues in the C-terminal α-helix by Ala destabilizes the tertiary structure considerably and that the effects of replacing Tyr20 or Tyr27, which are forming direct contacts with Pro residues from the N-terminal segment, are much larger than for Tyr21.

Our study additionally revealed that the turn region encompassing residues 10 to 14 is very sensitive to amino acid replacements. Whereas mutating Pro14 in PYY to Ala did not change the rigidity of the helical hairpin significantly, shifting Pro from position 14 to 13 resulted in formation of a non-back-folded species. Therefore it can be safely excluded that Pro14 forces the backbone to adopt a PP-type fold. In contrast, a shift of a single Pro residue by one position in sequence seems to be sufficient to convert peptides that adopt the PP-fold into structures that more closely resemble the conformation of NPY. NPY is clearly not back-folded, although the Pro residues at positions 2, 5 and 8 and Tyr at positions 20 and 27, which together form the hydrophobic core of the hairpin in PYY, are conserved. This points toward a crucial role for the residues of the β-turn region. Interestingly, replacing Ser13 by Ala results in a similar and even stronger destabilization of the helical hairpin. Pro residues are conserved at positions 13 and 14 in NPY and PYY, respectively, and Ser is highly conserved at position 13 in PYY. Computational studies indicate that hydrogen bonds departing from the hydroxyl group of Ser may help to stabilize the β-turn. These observations indicate that residues 13 and 14 serve as a hinge region, and that conformational preferences in that segment are critical to enable stable buildup of the helical hairpin. Although the lifetimes of particular conformational states may be too short to allow characterization through meaningful proton-proton NOEs, our study suggests that backbone dynamics data can be employed for this task. The data reveal that despite significant destabilization of the helical hairpin in some mutants, transient contacts between the N-terminal segment and the helix are formed much more often than it would be expected for a freely diffusing peptide chain. We therefore believe that these mutants sample conformational space similar to what would be expected for a molten globular state, and that they can be considered as models of transient structures being formed during folding of wild-type PYY.

Our data indicate that formation of secondary structure in PYY occurs to some extent independently from buildup of tertiary structure. N-terminal truncation mutants of NPY possess a helix albeit at somewhat reduced stability [38]. Moreover, NPY possesses no tertiary structure, but is clearly helical, and all the non back-folded mutants of PYY are still helical. It is therefore reasonable to assume that formation of the C-terminal α-helix in NPY occurs independently and that the tertiary contacts are initiated by hydrophobic contacts made with a pre-formed helix. Once the C-terminal helix is established, however, the PP-fold is formed cooperatively, because contacts of *all* Pro residues in the segment 2 to 8 as well as both Tyr20 and Tyr27 are required. The data suggest that the final conformation is stabilized by a manifold of weak interactions, all of which are required to drive folding to the correct state. This view is strongly supported by the thermal denaturation experiments with PYY and Tyr7-PYY and the measurements of the latter in water-methanol mixtures.

In summary, our results on the folding of an unconstrained helical hairpin suggest that both a turn-promoting sequence and specific non-local Pro-Tyr interactions add to the overall stability of the native fold. While a disfavorable sequence in the β-turn region suffices to destabilize the PP-fold significantly, a favorable sequence alone does not enforce it. Overall, the system should therefore prove useful to investigate the individual roles of the turn and the hydrophobic cluster in the folding dynamics of helical hairpins, for example using ϕ-value analysis [6]. Such a study has recently been conducted for a 16-residue β-hairpin from which it was suggested that turn formation is the rate-limiting step during folding, whereas the hydrophobic cluster between the strands slows unfolding. In remarkable analogy to the helical hairpin studied by us it was proposed that main chain and side chain residues of the turn region of this β-hairpin form a characteristic hydrogen bond network, which might play a critical role in stabilizing the folding transition state[39]. While we have drawn a comprehensive picture of the molecular interactions underlying the thermodynamics of helical hairpin formation in PYY, studies of the folding kinetics of PYY and the mutants presented here are now needed to shed more light on the mechanistic details of helical hairpin formation.

5.3 Conclusions

By NMR spectroscopy and denaturation experiments on a series of mutants we could demonstrate that tertiary structure formation in PYY is strictly dependent on the presence of a number of specific contacts. We could also demonstrate that formation of the PP-fold from preformed secondary structures occurs cooperatively. Most of the Pro->Ala or the Tyr->Ala mutants possess increased backbone dynamics, and the differences in N-terminal mobility among them reflect various degrees to which they sample conformations close to the PP-fold. Molten globules are generally considered being structurally related to the native structure with side-chain conformations less well defined, but many native-like tertiary contacts at least being transiently formed. In that respect many of the mutants studied in this work may actually be considered to be similar to the molten globular states of PYY.

The results may also have implications for our understanding of the binding of these peptides to their cognate membrane–embedded receptors. We have recently postulated that receptor binding is preceded by binding of the ligands to the membrane [22]. Changes in solvent properties accompany this change in environment. The solvent mixture experiments may therefore simulate the structural transition occurring when a peptide diffuses from the bulk solution towards the membrane. We have noticed that peptide structures are very similar when bound to either micelles or in methanol [22] and hence the transition from the membrane-bound state to water may be mimicked by the build-up of tertiary structure in PYY when changing from methanol to water solvent.

5.4 Materials and Methods

All peptides described in this paper have been produced by recombinant methods. They were expressed as insoluble fusions to ketosteroid isomerase, from which they were liberated through cyanobromide cleavage under denaturing conditions. Fusion peptides were expressed in M9 minimal media containing ^{15}N-NH_4Cl as the sole nitrogen source, and verified by their MS and [^{15}N,^{1}H]-HSQC spectra. C-terminal amidation was performed by enzymatic conversion of an extra Gly residue into an amide function using the α-amidating peptidyl glycine amidase (PAM).

The structures of pPYY, Tyr7-PYY and Ala14-PYY in solution were determined following established procedures using 2mM peptide samples at 28°C, pH 4.2 in 20mM deuterated acetate buffer on a Bruker AV-700. 15N relaxation data were recorded at 500 MHz proton frequency on 1mM uniformly 15N labeled samples under similar conditions. Procedures for the spectroscopic work were previously described in more detail [17, 21]. The structural characterization of the N-terminal polyproline helix proved challenging for two reasons: First, contributions to its stability arise both from medium-range contacts within the segment and long-range contacts to the C-terminal α-helix. Second, whilst a number of long-range proton-proton NOEs between protons of the C-terminal helix and the N-terminal segment clearly define the structure of the N terminus in the back-folded (native) state of PYY, the lack of both medium-range as well as long-range NOEs in the destabilized or denatured states results in great uncertainty under denaturing conditions or in destabilized mutants as to the presence of residual structure of the polyproline helix or transient back-folding to the C-terminal helix. We therefore chose to probe for residual structure by a combination of three different parameters, the $^{15}N\{^{1}H\}$-NOE, $^{3}J_{HN\alpha}$ scalar couplings, and residual dipolar couplings (RDCs) derived from samples that were partially aligned in stretched polyacrylamide gels. We have previously established a relationship between structure and backbone-dynamics that allowed us to quantify the extent of back-folding as an ensemble-averaged quantity of populations of fully back-folded and fully flexible N termini solely based on $^{15}N\{^{1}H\}$-NOE values [17, 21]. The fact that this model is consistent also with RDCs and $^{3}J_{HN\alpha}$ scalar couplings provides further evidence of its validity and usefulness for characterizing folding between two structurally well-characterized states by a single progress variable.

References

1. Cowley, D. J., Hoflack, J. M., Pelton, J. T. & Saudek, V. (1992) Structure of neuropeptide Y dimer in solution, *Eur. J. Biochem. 205*, 1099-106.

2. Sudol, M. (1996) Structure and function of the WW domain, *Prog. Biophys. Mol. Biol. 65*, 113-32.

3. Munoz, V., Thompson, P. A., Hofrichter, J. & Eaton, W. A. (1997) Folding dynamics and mechanism of beta-hairpin formation, *Nature. 390*, 196-9.

4. Ramirez-Alvarado, M., Blanco, F. J. & Serrano, L. (1996) De novo design and structural analysis of a model beta-hairpin peptide system, *Nat. Struct. Biol. 3*, 604-12.

5. Hughes, R. M. & Waters, M. L. (2006) Model systems for beta-hairpins and beta-sheets, *Curr. Opin. Struct. Biol. 16*, 514-24.

6. Fersht, A. (1999) *Structure and Mechanism in Protein Science*, W.H. Freeman, New York.

7. de Alba, E., Jimenez, M. A. & Rico, M. (1997) Turn residue sequence determines beta-hairpin conformation in designed peptides, *J. Am. Chem. Soc. 119*, 175-183.

8. Haque, T. S. & Gellman, S. H. (1997) Insights on beta-hairpin stability in aqueous solution from peptides with enforced type I' and type II' -turns, *J. Am. Chem. Soc. . 119*, 2303-2304.

9. Phillips, S. T., Piersanti, G. & Bartlett, P. A. (2005) Quantifying amino acid conformational preferences and side-chain-side-chain interactions in beta-hairpins, *Proc. Natl. Acad. Sci. U S A. 102*, 13737-42.

10. Kuroda, Y., Nakai, T. & Ohkubo, T. (1994) Solution structure of a de novo helical protein by 2D-NMR spectroscopy, *J. Mol. Biol. 236*, 862-8.

11. Du, D. & Gai, F. (2006) Understanding the folding mechanism of an alpha-helical hairpin, *Biochemistry. 45*, 13131-9.

12. Braisted, A. C. & Wells, J. A. (1996) Minimizing a binding domain from protein A, *Proc. Natl. Acad. Sci. U S A. 93*, 5688-92.

13. Betz, S. F., Bryson, J. W. & DeGrado, W. F. (1995) Native-like and structurally characterized designed alpha-helical bundles, *Curr. Opin. Struct. Biol. 5*, 457-63.

14. Fezoui, Y., Weaver, D. L. & Osterhout, J. J. (1994) De novo design and structural characterization of an alpha-helical hairpin peptide: a model system for the study of protein folding intermediates, *Proc. Natl. Acad. Sci. U S A. 91*, 3675-9.

15. Blundell, T. L., Pitts, J. E., Tickle, S. P. & Wu, C. W. (1981) X-ray analysis (1.4 A resolution) of avian pancreatic polypeptide: small globular protein hormone, *Proc. Natl. Acad. Sci. U S A. 78*, 4175-9.

16. Keire, D. A., Kobayashi, M., Solomon, T. E. & Reeve, J. R., Jr. (2000) Solution structure of monomeric peptide YY supports the functional significance of the PP-fold, *Biochemistry. 39*, 9935-42.

17. Lerch, M., Mayrhofer, M. & Zerbe, O. (2004) Structural similarities of micelle-bound peptide YY (PYY) and neuropeptide Y (NPY) are related to their affinity profiles at the Y receptors, *J. Mol. Biol. 339*, 1153-68.

18. Monks, S. A., Karagianis, G., Howlett, G. J. & Norton, R. S. (1996) Solution structure of human neuropeptide Y, *J. Biomol. NMR. 8*, 379-90.

19. Gellman, S. H. & Woolfson, D. N. (2002) Mini-proteins Trp the light fantastic, *Nat. Struct. Biol. 9*, 408-10.
20. Bhattacharyya, R. & Chakrabarti, P. (2003) Stereospecific interactions of proline residues in protein structures and complexes, *J. Mol. Biol. 331*, 925-40.
21. Bader, R., Bettio, A., Beck-Sickinger, A. G. & Zerbe, O. (2001) Structure and Dynamics of Micelle-bound Neuropeptide Y: Comparison with unligated NPY and Implications for Receptor Selection, *J. Mol. Biol. 305*, 307-392.
22. Bader, R. & Zerbe, O. (2005) Are hormones from the neuropeptide Y family recognized by their receptors from the membrane-bound state?, *ChemBioChem. 6*, 1520-34.
23. Woll, M. G. & Gellman, S. H. (2004) Backbone thioester exchange: a new approach to evaluating higher order structural stability in polypeptides, *J Am Chem Soc. 126*, 11172-4.
24. Lerch, M., Gafner, V., Bader, R., Christen, B., Folkers, G. & Zerbe, O. (2002) Bovine pancreatic polypeptide (bPP) undergoes significant changes in conformation and dynamics upon binding to DPC micelles, *J. Mol. Biol. 322*, 1117-33.
25. Chang, P. J., Noelken, M. E. & Kimmel, J. R. (1980) Reversible dimerization of avian pancreatic polypeptide, *Biochemistry. 19*, 1844-9.
26. Bianchi, E., Rampone, R., Tealdi, A. & Ciferri, A. (1970) The role of aliphatic alcohols on the stability of collagen and tropocollagen, *J. Biol. Chem. 245*, 3341-5.
27. Herskovits, T. T., Gadegbeku, B. & Jaillet, H. (1970) On the structural stability and solvent denaturation of proteins. I. Denaturation by the alcohols and glycols, *J. Biol. Chem. 245*, 2588-98.
28. Kamatari, Y. O., Konno, T., Kataoka, M. & Akasaka, K. (1996) The methanol-induced globular and expanded denatured states of cytochrome c: a study by CD fluorescence, NMR and small-angle X-ray scattering, *J. Mol. Biol. 259*, 512-23.
29. de Jongh, H. H., Killian, J. A. & de Kruijff, B. (1992) A water-lipid interface induces a highly dynamic folded state in apocytochrome c and cytochrome c, which may represent a common folding intermediate, *Biochemistry. 31*, 1636-43.
30. Babu, K. R., Moradian, A. & Douglas, D. J. (2001) The methanol-induced conformational transitions of beta-lactoglobulin, cytochrome c, and ubiquitin at low pH: a study by electrospray ionization mass spectrometry, *J. Am. Soc. Mass. Spectrom. 12*, 317-28.
31. Wang, Y.-F., Ho, M.-Y. & Ho, Y.-P. (2004) Using mass spectrometry to probe subtle differences in conformations of several cytochromes c in aqueous and methanol solutions, *J. Mass Spectrom. 39*, 1523-1530.
32. Bychkova, V. E., Dujsekina, A. E., Klenin, S. I., Tiktopulo, E. I., Uversky, V. N. & Ptitsyn, O. B. (1996) Molten globule-like state of cytochrome c under conditions simulating those near the membrane surface, *Biochemistry. 35*, 6058-63.
33. Alonso, D. O. & Daggett, V. (1995) Molecular dynamics simulations of protein unfolding and limited refolding: characterization of partially unfolded states of ubiquitin in 60% methanol and in water, *J. Mol. Biol. 247*, 501-20.
34. Santoro, M. M. & Bolen, D. W. (1988) Unfolding free energy changes determined by the linear extrapolation method. 1. Unfolding of phenylmethanesulfonyl alpha-chymotrypsin using different denaturants, *Biochemistry. 27*, 8063-8.

35. Ding, F., Buldyrev, S. V. & Dokholyan, N. V. (2005) Folding Trp-cage to NMR resolution native structure using a coarse-grained protein model, *Biophys. J. 88*, 147-55.

36. Neidigh, J. W., Fesinmeyer, R. M. & Andersen, N. H. (2002) Designing a 20-residue protein, *Nat. Struct. Biol. 9*, 425-30.

37. Yu, H., Rosen, M. K., Shin, T. B., Seidel-Dugan, C., Brugge, J. S. & Schreiber, S. L. (1992) Solution structure of the SH3 domain of Src and identification of its ligand-binding site, *Science. 258*, 1665-8.

38. Arvidsson, K., Jarvet, J., Allard, P. & Ehrenberg, A. (1994) Solution structure by 1H and dynamics by natural abundance 13C NMR of a receptor recognising peptide derived from a C-terminal fragment of neuropeptide Y, *J. Biomol. NMR. 4*, 653-72.

39. Du, D., Tucker, M. J. & Gai, F. (2006) Understanding the Mechanism of b-Hairpin Folding via F-Value Analysis, *Biochemistry. 45*, 2668-2678.

5.5 Supplementary material

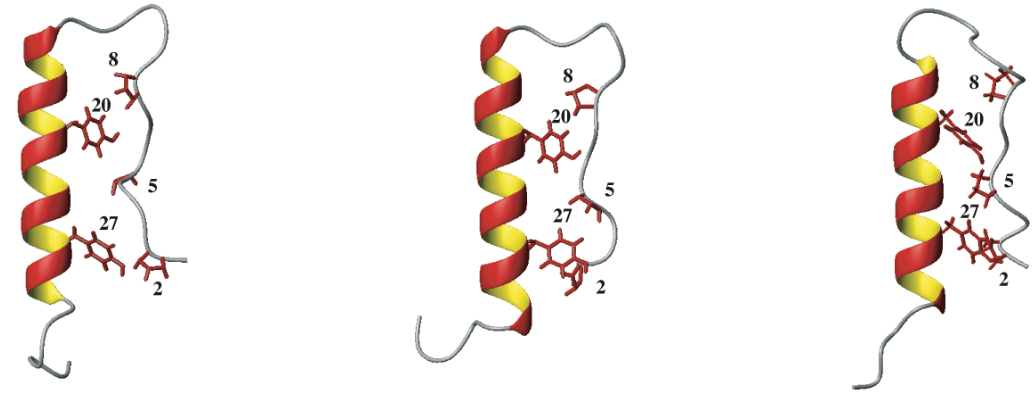


Figure S1: Representative structures of Ala14-pPYY (left), pPYY (middle) and bPP (right).

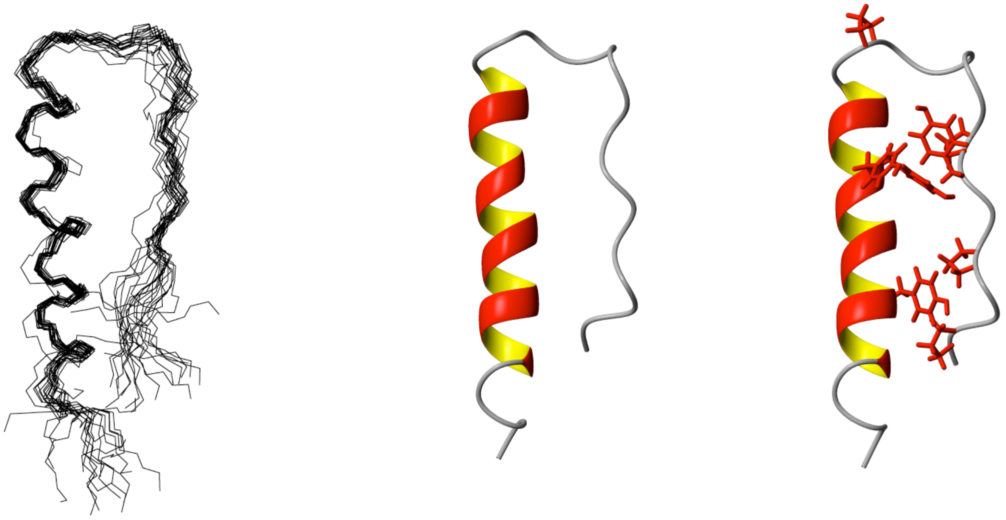


Figure S2: Structure of Tyr7-PYY as determined at pH 4.1, 301K: Backbone presentation of the superposition of single 20 lowest energy conformers (left), and single conformer (middle) with sidechains of Pro and Tyr residues depicted (right).

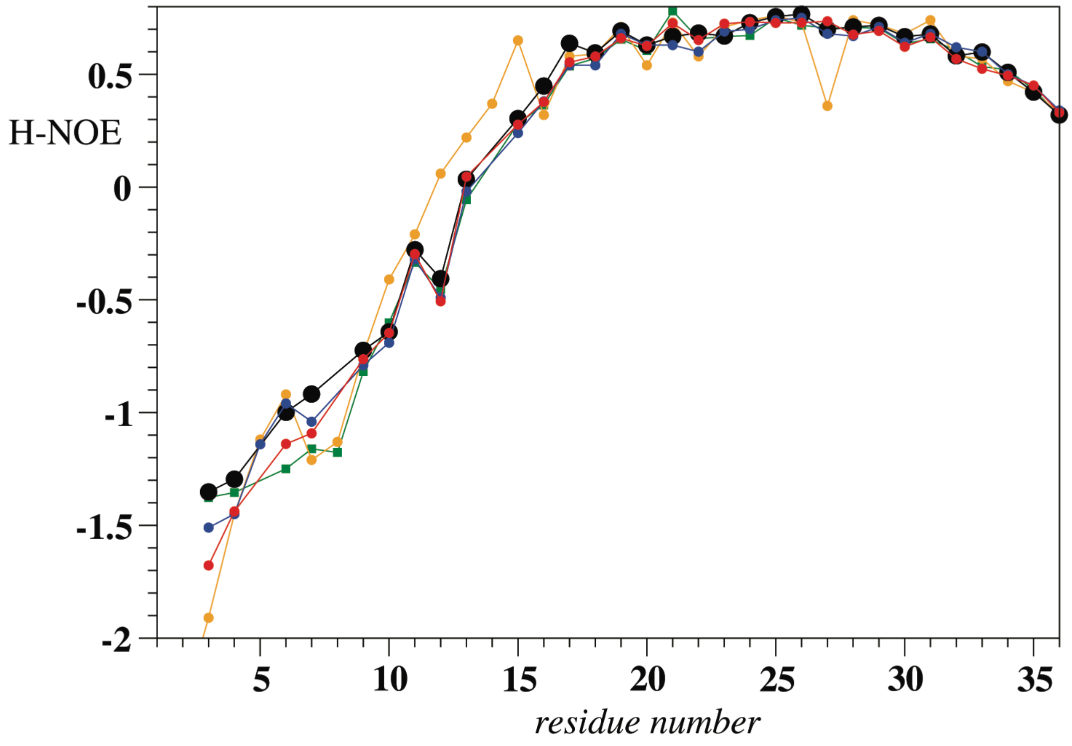


Figure S3: $^{15}N\{^{1}H\}$-NOE of PYY, Ala2-PYY, Ala5-PYY, Ala8-PYY and Ala2,5,8,14-PYY at pH 6.0 in 20 mM MES buffer, 300mM DPC using the same color coding and symbol as in Fig. 2.

Solvent mixture (methanol-water) experiments for Tyr7-PYY:

Mixtures of water and methanol may be particularly useful to study helical hairpin formation in PYY. We have followed the amide proton chemical shifts *vs*. the methanol content. The data for selected residues of Tyr7-PYY, for which the transition between non-back-folded and back-folded species occurs at a larger methanol-to-water ratio compared to PYY and hence is better visible. i) Chemical shifts of residues 3,4,6,7,9,10, 12 17, 18 (23) and 24 of Tyr7-PYY, which are all part of the back-fold interface and hence experience a major structural transition, display a characteristic sigmoidal curve. ii) An approximately linear change occurs for a amide protons being part of an unstructured segment, for which the major change is reflected in the decrease of solvent-coordination upon increasing the content of methanol, e.g. for residues 35 and 36 of Tyr7-PYY as well as from residues 6,7 11, 35 and 36 of PYY. The resonance frequencies of these protons shift to higher field with increasing methanol content. iii) Amide proton resonances of residues 20, 22, 25, 26, 29 and 30-32, all of which are pointing away from the back-folding interface, are shifted to lower field with increasing methanol content. These changes can be attributed to the fact that the isolated helix and hence the involved hydrogen bonds are stabilized in methanol.

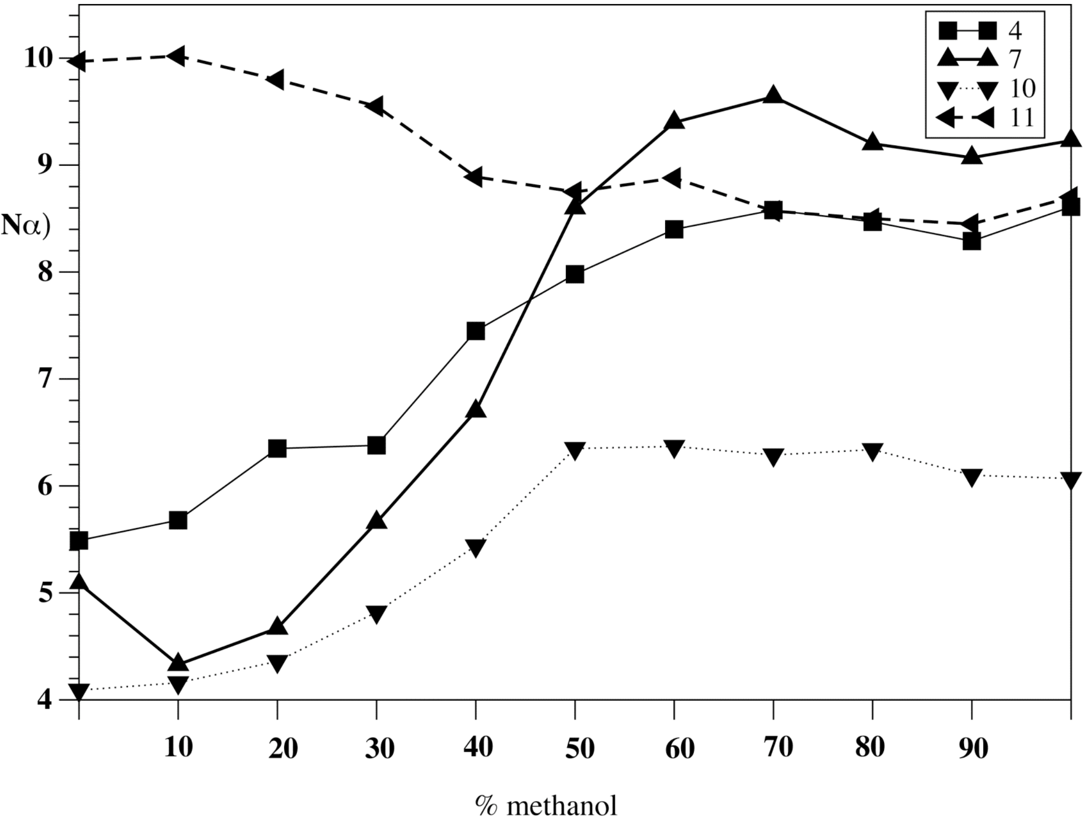


Figure S4: ^{3}J(HNα) scalar coupling constants for Tyr7-PYY in various water-methanol mixtures.

The data display sigmoidal curves. Interestingly, while the H-NOE data indicate that the N-terminal segment is largely flexible at 100% methanol, the scalar coupling data are only compatible with incomplete averaging of dihedral angles. While further work in our laboratory is in progress to investigate this issue in more detail, we speculate that the presence of Pro residues restricts conformational space for backbone rotameric states.

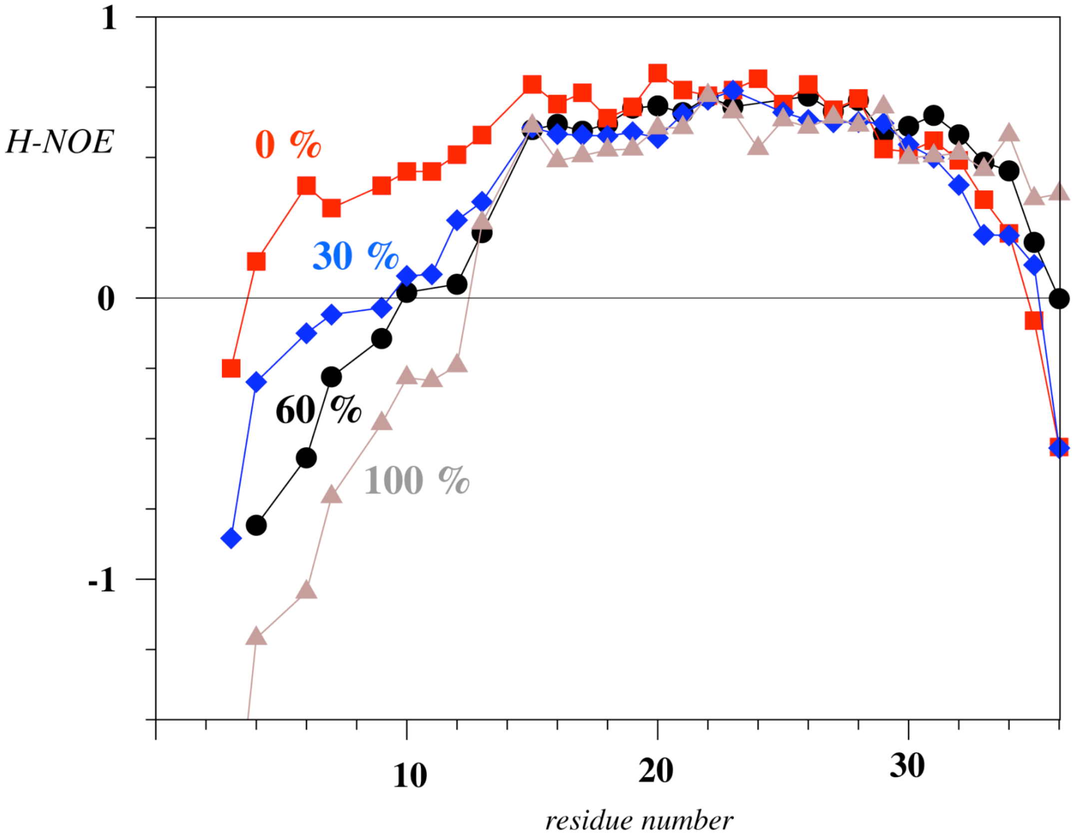


Figure S5: H-NOE data of PYY in various water-methanol mixtures (percentage methanol indicated).

Dimerization of PYY and Tyr7-PYY

It should be noted that the related avian polypeptide (aPP) exists in dimeric form in the crystal structure. Therein, the interface between the monomeric subunits is formed by a hydrophobic cluster, in which the π-systems of Tyr7, Tyr21 and Phe20 stack onto each other. In PYY position 7 is occupied by Ala, and hence we decided to replace it by Tyr. The solution structure of Tyr7-PYY and the $^{15}N\{^1H\}$-NOE data indicate that the backbone of Tyr7-PYY is very similar to PYY, and slightly less rigid in the turn region comprising residues 10 to 14. The K_d for homodimerization of PYY is reported to be 22 mM at pH 5. It is difficult to measure K_d's in that range due to the limited solubility of the peptide at very high concentrations, but dilution NMR experiments conducted with PYY suggest a similar value at pH 4.1 (35 ±18 mM). It is also difficult to follow changes of backbone dynamics at very low concentrations to derive estimates of the contribution of dimerization to the stability of the back-fold. However, it is still possible to monitor chemical shift changes upon dilution. The amide proton frequency of Ala12 is very sensitive to the extent of back-folding: While it is at about 7.3 ppm in PYY, it is measured to be 7.9 ppm in the strongly destabilized mutant Ala8. The amide proton shift of Ala12 shifts by -0.06 ppm upon dilution from 1mM to 20uM. At the latter concentration the population of the monomer is higher than 99%, while it is between 91% and 95% at 1mM for K_d's in the range between 20-40mM. The data indicate that the dimer contributes to the stability of the helical hairpin only to a small extent, so that the gross effects must be due to changes within the monomer. We therefore feel justified to treat the system as a monomer to simplify the analysis.

Turn geometry optimizations and energy calculations

The turn segments of aPP and PYY comprising residues 11 – 15 were geometry optimized both in fully extended and in their native conformations. For the optimization of the native turn conformation, the coordinates of the backbone heavy atoms of residues 11 and 15 were held fixed in the positions as initially derived from the crystal structure of aPP (1) and the NMR structure of PYY (2), respectively. Geometry optimizations were performed using the B3LYP method (3) and the lacvp double-ζ quality basis set, as implemented in the Jaguar 4.2 program (4). Polarization effects from the surrounding solvent were taken into account by using a dielectric cavity model with a dielectric constant of ε=78.8 and a probe radius of 1.4. Table S2 summarizes the electronic energies after optimization of aPP and PYY in extended and turn conformation and the energy difference associated with the conformational change from extended to turn structures.

Table S1

Molecule AA seq 11-15	**conformation**	**el. energy [Hartree]**	**$\Delta E_{ext\text{->}turn}$ [kcal/mole]** (error is on the order of 5 kcal/mole)
aPP	extended	-1922,551967	
DAPYE	turn	-1922,561057	-5,70
PYY	extended	-1919,149773	
DASPE	turn	-1919,137608	7,63

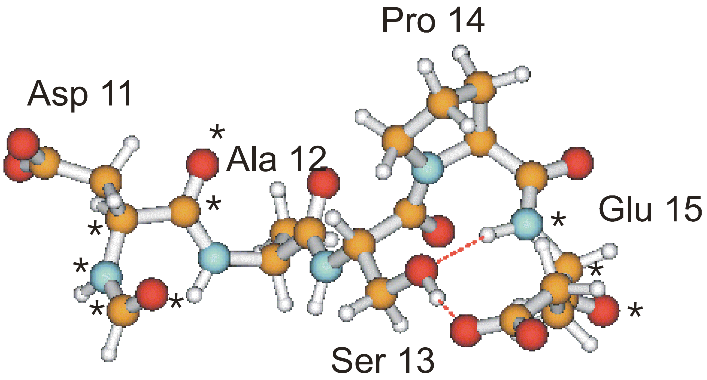


Figure S6: Optimized turn conformation of PYY. Atoms held fixed during the optimization are labeled with a star.

Turn formation in aPP is predicted to be slightly favourable, by contrast to turn formation in PYY, for which calculations predict a slightly unfavourable energy. As mentioned in the manuscript text, turn formation in PYY is not consistent with a Pro in position 13 (by contrast to aPP), nor does it tolerate substitution of Ser13 by Ala. Together, these observations point towards a critical role played by Ser13 in turn formation of PYY, rather than a geometry constraining function exerted by Pro14. One possible mechanism as to how Ser13 may stabilize the turn is evident from Figure S4, showing two hydrogen bonds from and to the side-chain hydroxyl group of Ser13. One is between the hydrogen of the OH group and the side-chain carboxyl group of Glu15 (d=1.56 Å), whilst the other involves the hydroxyl oxygen and the backbone amide group of Glu15 (d=1.91 Å). Substitution of Ser13 by Ala may therefore remove two turn-supporting interactions and hence critically destabilize the back-folded structure.

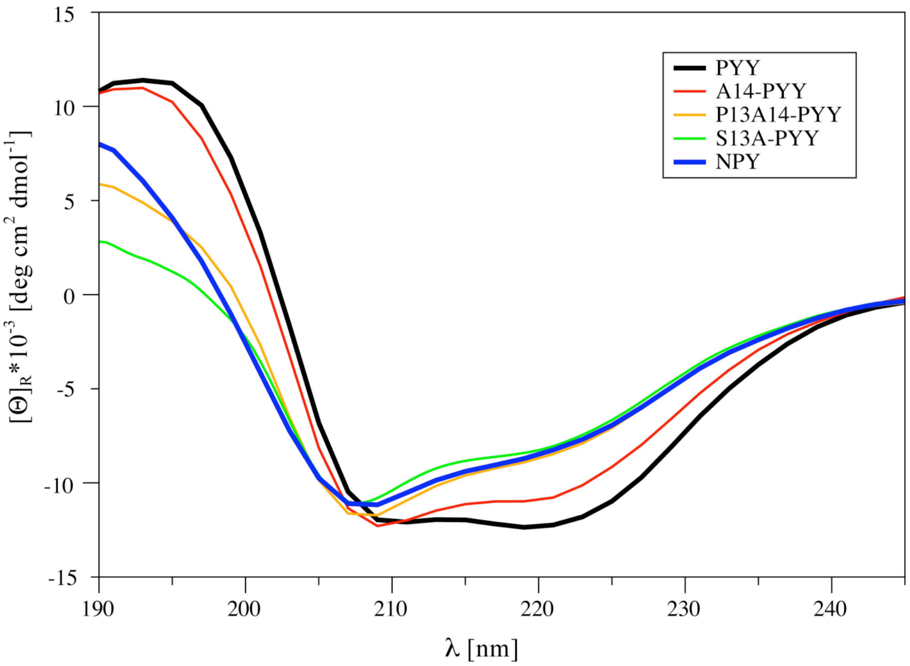


Figure S7: CD spectra of pPYY mutants (α-hairpin turn region), pPYY and NPY. Spectra were recorded in 1mM Na-acetate buffer (pH 4.4) at 298K.

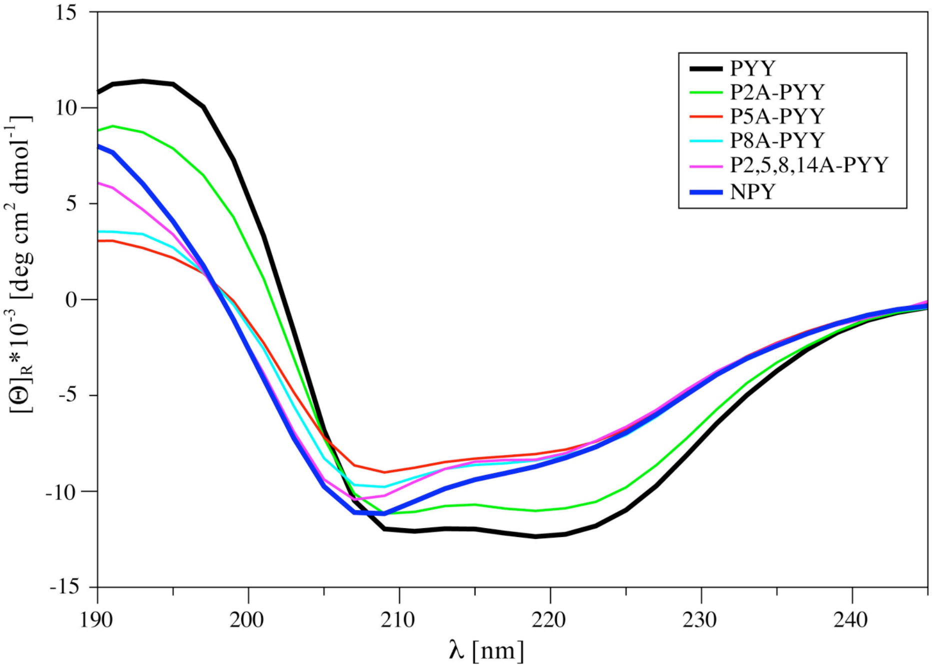


Figure S8: CD spectra of pPYY Pro->Ala mutants (N-terminal segment), pPYY and NPY. Spectra were recorded in 1mM Na-acetate buffer (pH 4.4) at 298K.

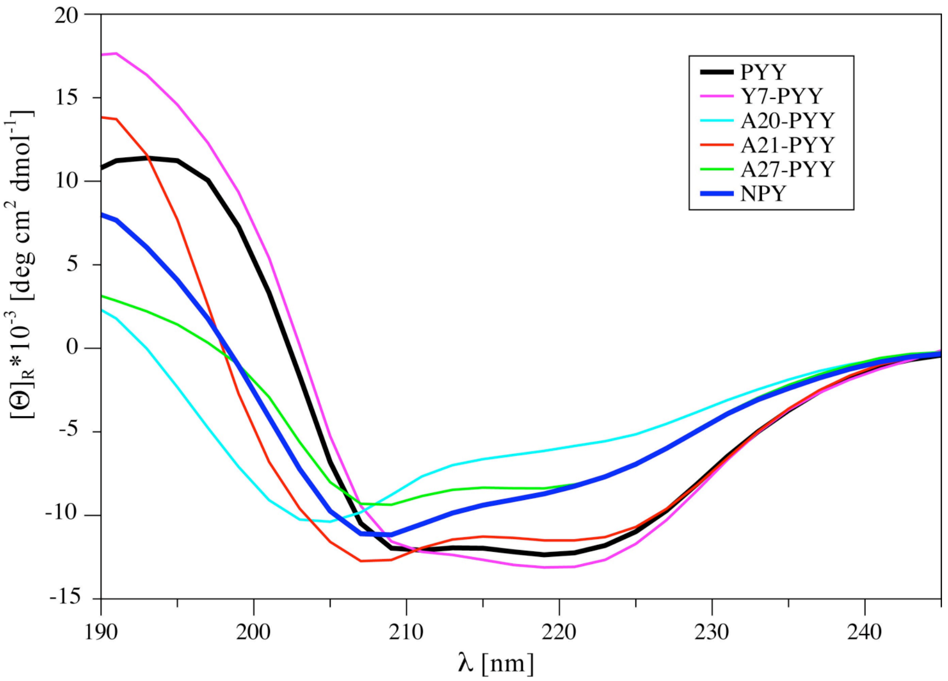


Figure S9: CD spectra of pPYY, Y7-pPYY, pPYY single Tyr->Ala mutants and NPY. Spectra were recorded in 1mM Na-acetate buffer (pH 4.4) at 298K.

Table S2: Chemical shifts for 2mM P14A-pPYY in 90% $H_2O/10\%^2H_2O$, pH 4.1, T 301K, referenced to the signal of residual HDO at 4.72 ppm.

N	Res.	H^N	H'^{α}	H^{β}	others
1	TYR	-	4.43	2.97, 3.17	δH 7.18, 7.18; εH 6.84, 6.84; ηOH -
2	PRO		4.47	1.90, 2.29	γCH_2 1.91, 1.91; δCH_2 3.21, 3.65
3	ALA	8.38	4.32	1.33	
4	LYS	8.26	3.68	1.40, 1.57	γCH_2 1.08, 1.08; δCH_2 1.25, 1.25; εCH_2 2.91, 2.91; ζNH_3^+ -
5	PRO		4.34	1.85, 2.27	γCH_2 1.77, 1.77; δCH_2 3.19, 3.50
6	GLU	8.45	4.22	1.85, 1.93	γCH_2 2.36, 2.36; εH -
7	ALA	8.31	3.69	1.11	
8	PRO		4.31	1.77, 2.08	γCH_2 1.61, 1.61; δCH_2 2.93, 2.93-
9	GLY	8.22	3.85, 3.97		
10	GLU	8.34	4.09	1.98, 1.98	γCH_2 2.37, 2.37; εH -
11	ASP	8.51	4.62	2.74, 2.79	δH -
12	ALA	7.46	4.20	1.32	
13	SER	8.35	4.40	4.01, 4.30	γOH -
14	ALA	8.76	4.09	1.46	

15	GLU	8.49	4.11	1.97, 2.08	γCH$_2$ 2.39, 2.39; εH -
16	GLU	7.87	3.95	1.97, 2.23	γCH$_2$ 2.32, 2.39; εH -
17	LEU	8.40	3.97	1.56, 1.82	γH 1.32; δCH$_3$ 0.81, 0.81
18	SER	8.28	4.26	4.02, 4.02	γOH -
19	ARG	7.95	4.11	1.81, 1.91	γCH$_2$ 1.60, 1.60; δCH$_2$ 3.18, 3.18; εNH 7.34; ηNH$_2$ -, -
20	TYR	8.05	4.33	2.95, 3.12	δH 6.65, 6.65; εH 6.47, 6.47; ηOH -
21	TYR	8.64	3.96	2.94, 3.07	δH 7.05, 7.05; εH 6.80, 6.80; ηOH -
22	ALA	7.98	4.11	1.51	
23	SER	8.17	4.29	3.93, 4.03	γOH -
24	LEU	8.60	4.00	1.60, 1.60	γH 1.17; δCH$_3$ 0.82, 0.82
25	ARG	8.08	3.92	1.86, 1.86	γCH$_2$ 1.55, 1.71; δCH$_2$ 3.14, 3.20; εNH 7.24; ηNH$_2$ -, -
26	HIS	7.87	4.41	3.29, 3.36	δ^{1}NH -; δ^{2}H 7.27; ε^{1}H 8.62; ε^{2}NH -
27	TYR	8.18	4.09	3.14, 3.14	δH 6.96, 6.96; εH 6.69, 6.69; ηOH -
28	LEU	8.55	3.86	1.77, 1.77	γH 1.43; δCH$_3$ 0.82, 0.82
29	ASN	7.96	4.41	2.78, 2.83	δNH$_2$ 6.86, 7.55
30	LEU	7.66	4.07	1.40, 1.64	γH 1.48; δCH$_3$ 0.73, 0.73
31	VAL	7.84	3.82	1.92	γCH$_3$ 0.62, 0.62
32	THR	7.74	4.23	4.15	γCH$_3$ 1.18; γOH -
33	ARG	7.84	4.20	1.82, 1.82	γCH$_2$ 1.62, 1.62; δCH$_2$ 3.13, 3.13; εNH 7.15; ηNH$_2$ -, -
34	GLN	8.07	4.16	1.86, 1.95	γCH$_2$ 2.28, 2.28; εNH$_2$ -, -
35	ARG	8.10	4.15	1.63, 1.63	γCH$_2$ 1.36, 1.44; δCH$_2$ 3.06, 3.06; εNH 7.09; ηNH$_2$ -, -
36	TYR	8.01	4.52	2.85, 3.07	δH 7.09, 7.09; εH 6.76, 6.76; ηOH -

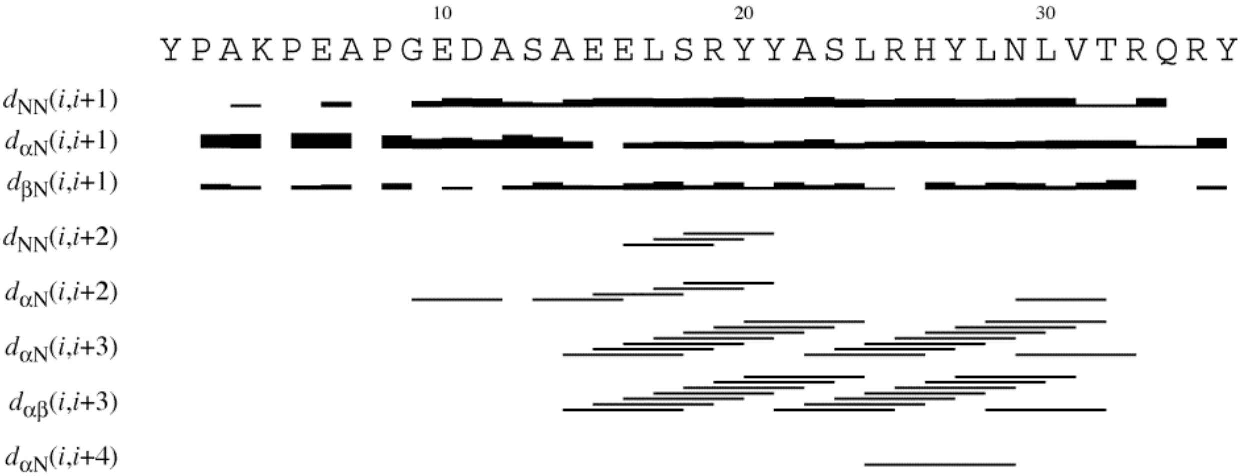


Figure S10: Sequence plot of restraints used during the structure calculation of P14A-pPYY.

Table S3: Information on the structure calculation of P14A-pPYY in solution.

Distance restraints	Total	342
	Intra-residual	84
	Sequential (*i* - *j* = 1)	140
	Medium (*i* - *j* = 2, 3, 4)	99
	Long-range	19
Dihedral angle restraints		147
RMSD (Å)[a]		
	Pro5-Val31 all heavy atoms	2.48 ± 0.80
	Leu14-Arg29 backbone	0.59 ± 0.30
	Leu14-Arg29 all heavy atoms	1.46 ± 0.29
Structure check[b] (Average %)		
	Tyr1-Tyr36	98.8 (1.2)
NOE constraint violations	Number > 0.1 Å	1
	Maximum (Å)	0.21
Dihedral angle constraint violations	Number > 2.5 degrees	0
AMBER energies (kcal/mol)[c]	Total	-581.8
	Van der Waals	-31.8
	Electrostatic	-875.1
[a] N, C_α, C´ atoms.		
[b] Percentage of the ϕ, φ angles falling into the most favored (disallowed) regions within the allowed Ramachandran regions for the 20 refined structures.		
[c] AMBER energies are given as the sum of solute-solute and solute-water interactions.		

Table S4: Chemical shifts for 1mM A7Y-pPYY in 90% H_2O/10%^2H_2O, pH 4.1, T 301K, referenced to the signal of residual HDO at 4.72 ppm.

N	Res.	H^N	H$^{'a}$	H^b	others
1	TYR	-	4.46	3.01, 3.19	δH 7.22, 7.22; εH 6.89, 6.89; ηOH -
2	PRO		4.51	1.98, 2.32	γCH$_2$ 1.91, 1.91; δCH$_2$ 3.20, 3.67
3	ALA	8.38	4.37	1.35	
4	LYS	8.27	3.48	1.35, 1.52	γCH$_2$ 0.94, 1.22; δCH$_2$ 1.59, 1.59; εCH$_2$ 2.93, 2.93; ζNH$_3^+$ -
5	PRO		4.39	1.88, 2.29	γCH$_2$ 1.70, 1.73; δCH$_2$ 3.08, 3.45
6	GLU	8.54	4.34	1.87, 1.91	γCH$_2$ 2.30, 2.36; εH -
7	TYR	8.6	3.7	2.62, 2.72	δH 6.63, 6.63; εH 6.61, 6.61; ηOH -
8	PRO		4.24	1.70, 2.04	γCH$_2$ 1.26, 1.51; δCH$_2$ 2.37, 3.13
9	GLY	5.45	3.56, 4.25		
10	GLU	8.44	4.17	2.06, 2.06	γCH$_2$ 2.48, 2.48; εH -
11	ASP	8.41	4.73	2.72, 2.87	δH -
12	ALA	7.09	4.16	1.16	
13	SER	8.32	4.68	4.02, 4.39	γOH -
14	PRO		4.28	1.98, 2.41	γCH$_2$ 2.08, 2.22; δCH$_2$ 3.91, 3.96
15	GLU	8.51	4.15	1.97, 2.11	γCH$_2$ 2.40, 2.44; εH -
16	GLU	7.85	4.01	1.90, 2.29	γCH$_2$ 2.36, 2.41; εH -
17	LEU	8.51	4.02	1.55, 1.64	γH 1.87; δCH$_3$ 0.80, 0.89
18	SER	8.48	4.36	4.05, 4.05	γOH -
19	ARG	7.88	4.16	1.95, 1.95	γCH$_2$ 1.64, 1.87; δCH$_2$ 3.18, 3.25; εNH 7.35; ηNH$_2$ -, -
20	TYR	8.01	4.30	3.08, 3.08	δH 6.65, 6.65; εH 6.56, 6.56; ηOH -
21	TYR	8.73	3.96	2.99, 3.18	δH 7.19, 7.19; εH 6.89, 6.89; ηOH -
22	ALA	7.98	4.16	1.55	-
23	SER	8.24	4.32	3.93, 4.05	γOH -
24	LEU	8.88	3.99	1.03, 1.71	γH 1.54; δCH$_3$ 0.91, 0.98
25	ARG	8.07	3.91	1.75, 1.92	γCH$_2$ 1.59, 1.59; δCH$_2$ 3.18, 3.24; εNH 7.20; ηNH$_2$ -, -
26	HIS	7.75	4.43	3.34, 3.40	δ^{1}NH -; δ^{2}H 7.37; ε^{1}H 8.65; ε^{2}NH -
27	TYR	8.24	4.09	3.20, 3.20	δH 6.99, 6.99; εH 6.71, 6.71; ηOH -
28	LEU	8.65	3.87	1.41, 1.77	γH 1.88; δCH$_3$ 0.83, 0.90

29	ASN	7.91	4.44	2.80, 2.87	δNH$_2$ 6.86, 7.52
30	LEU	7.64	4.11	1.41, 1.67	γH 1.52; δCH$_3$ 0.75, 0.75
31	VAL	7.8	3.87	1.95	γCH$_3$ 0.61, 0.64
32	THR	7.72	4.20	4.24	γCH$_3$ 1.22; γOH -
33	ARG	7.89	4.24	1.80, 1.88	γCH$_2$ 1.64, 1.64; δCH$_2$ 3.15, 3.15; εNH 7.14; ηNH$_2$ -, -
34	GLN	8.10	4.20	1.95, 2.02	γCH$_2$ 2.31, 2.31; εNH$_2$ -, -
35	ARG	8.12	4.18	1.67, 1.67	γCH$_2$ 1.40, 1.47; δCH$_2$ 3.10, 3.10; εNH 7.09; ηNH$_2$ -, -
36	TYR	8.04	4.55	2.88, 3.10	δH 7.13, 7.13; εH 6.80, 6.80; ηOH -

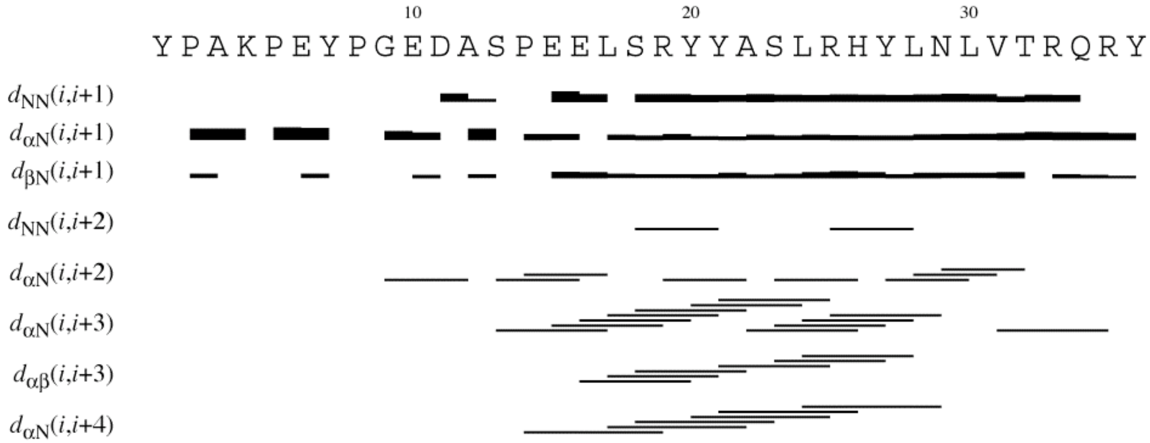


Figure S11: Sequence plot for restraints used during the structure calculation of A7Y-PYY.

Table S5: Information on the structure calculation of A7Y-pPYY in solution.

Distance restraints	Total	552
	Intra-residual	121
	Sequential ($i - j = 1$)	161
	Medium ($i - j = 2, 3, 4$)	152
	Long-range	75
Dihedral angle restraints		
RMSD (Å)[a]		
	Pro5-Val31 backbone	1.29 ± 0.44
	Pro5-Val31 all heavy atoms	2.21 ± 0.48
	Pro14-Val31 backbone	0.81 ± 0.30
	Pro14-Val31 all heavy atoms	1.72 ± 0.37
Structure check[b] (Average %)		
	Tyr1-Tyr36	99.2 (0.8)
NOE constraint violations	Number > 0.1 Å	0
	Maximum (Å)	-
Dihedral angle constraint violations	Number > 2.5 degrees	0
AMBER energies (kcal/mol)[c]	Total	-1112.0
	Van der Waals	239.1
	Electrostatic	-2561.1

[a] N, C_α, C´ atoms.

[b] Percentage of the ϕ, φ angles falling into the most favored (disallowed) regions within the allowed Ramachandran regions for the 20 refined structures.

[c] AMBER energies are given as the sum of solute-solute and solute-water interactions.

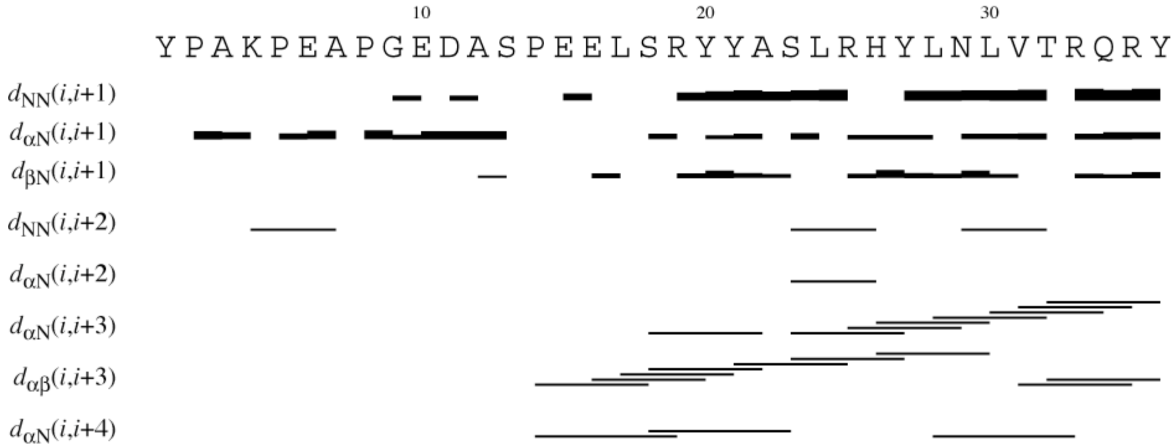


Figure S12: Sequence plot of restraints used during the structure calculation of pPYY in MeOH.

Table S6: Chemical Shifts of pPYY, 1mM, d^3-MeOH, 301K, referenced to residual CD$_2$**H**OH at 3.31 ppm.

N	Res.	H^N	H'^a	H^b	others
1	Tyr	-	-	-, -	δH -, -; εH -, -; ηOH -
2	Pro		4.43	1.97, 2.22	γCH$_2$ -, -; δCH$_2$ -, -
3	Ala	8.29	4.32	1.38	
4	Lys	8.16	4.63	1.86, 1.86	γCH$_2$ 1.69, 1.69; δCH$_2$ 1.51, 1.51; εCH$_2$ -, -; ζNH$_3^+$ -
5	Pro		4.49	-, -	γCH$_2$ 2.09, 2.09; δCH$_2$ 3.65, 3.80
6	Glu	8.29	4.36	1.94, 1.94	γCH$_2$ 2.11, 2.43; εH -
7	Ala	8.07	4.63	1.34	
8	Pro		4.35	1.97, 2.23	γCH$_2$ -, -; δCH$_2$ -, -
9	Gly	8.52	3.80, 3.92		
10	Glu	8.07	4.29	2.42, 2.42	γCH$_2$ 2.07, 2.15; εH -
11	Asp	8.27	4.67	2.79, 2.89	δH -
12	Ala	7.95	4.49	1.39	
13	Ser	8.13	4.7	3.98, 4.26	γOH -
14	Pro		4.3	1.94, 2.05	γCH$_2$ 1.69, 1.69; δCH$_2$ -, -
15	Glu	8.24	4.09	2.21, 2.21	γCH$_2$ 2.51, 2.59; εH -
16	Glu	7.96	4.11	2.17, 2.28	γCH$_2$ 2.51, 2.51; εH -
17	Leu	8.34	4.11	1.86, 1.86	γH 1.50; δCH$_3$ 0.93, 0.93
18	Ser	8.34	4.25	4.01, 4.10	γOH -

19	Arg	7.97	4.06	1.99, 1.99	$\gamma\mathrm{CH_2}$ 1.67, 1.67; $\delta\mathrm{CH_2}$ 3.21, 3.21; εNH 7.45; $\eta\mathrm{NH_2}$ -, -
20	Tyr	8.44	4.17	3.13, 3.18	δH 6.76, 6.76; εH 6.56, 6.56; ηOH -
21	Tyr	8.67	4.13	3.12, 3.12	δH 7.16, 7.16; εH 6.77, 6.77; ηOH -
22	Ala	8.58	4.05	1.57	
23	Ser	8.23	4.17	4.03, 4.06	γOH -
24	Leu	8.35	4.1	1.68, 1.68	γH 1.50; $\delta\mathrm{CH_3}$ 0.87, 0.87
25	Arg	8.49	3.96	1.99, 1.99	$\gamma\mathrm{CH_2}$ -, -; $\delta\mathrm{CH_2}$ 3.20, 3.20; εNH 7.44; $\eta\mathrm{NH_2}$ -, -
26	His	8.4	4.48	3.41, 3.51	δ^1NH -; δ^2H 7.30; ε^1H 8.73; ε^2NH -
27	Tyr	8.49	4.07	3.21, 3.21	δH 6.97, 6.97; εH 6.64, 6.64; ηOH -
28	Leu	8.7	3.92	1.97, 1.97	γH 1.50; $\delta\mathrm{CH_3}$ 0.93, 0.93
29	Asn	8.55	4.38	2.75, 3.10	$\delta\mathrm{NH_2}$ 7.02, 7.79
30	Leu	8.01	3.97	1.74, 1.74	γH 1.60; $\delta\mathrm{CH_3}$ 0.84, 0.84
31	Val	8.27	3.65	1.98	$\gamma\mathrm{CH_3}$ 0.75, 0.84
32	Thr	8.03	3.87	4.36	$\gamma\mathrm{CH_3}$ 1.28; γOH -
33	Arg	8.03	4.06	1.98, 1.98	$\gamma\mathrm{CH_2}$ 1.86, 1.86; $\delta\mathrm{CH_2}$ 3.16, 3.16; εNH 7.44; $\eta\mathrm{NH_2}$ -, -
34	Gln	8.15	4.02	2.27, 2.27	$\gamma\mathrm{CH_2}$ 2.07, 2.52; $\varepsilon\mathrm{NH_2}$ -, -
35	Arg	7.89	4.07	1.65, 1.75	$\gamma\mathrm{CH_2}$ 1.43, 1.52; $\delta\mathrm{CH_2}$ 3.01, 3.07; εNH 7.25; $\eta\mathrm{NH_2}$ -, -
36	Tyr	7.77	4.52	2.84, 3.20	δH 7.22, 7.22; εH 6.67, 6.67; ηOH -

Table S7: Chemical shifts for pPYY mutants, 1 mM samples in 90% H_2O/10%^2H_2O, pH 4.1, T 301K, referenced to the signal of residual HDO at 4.72 ppm

N	P2A-pPYY			P5A-pPYY			P8A-pPYY			P2,5,8,14A-pPYY			13P14A-pPYY			S13A-pPYY		
	Res.	N, ppm	H^N, ppm	Res.	N, ppm	H^N, ppm	Res.	N, ppm	H^N, ppm	Res.	N, ppm	H^N, ppm	Res.	N, ppm	H^N, ppm	Res.	N, ppm	H^N, ppm
1	TYR			TYR			TYR			TYR			TYR			TYR		
2	ALA	126.73	8.27	PRO			PRO			ALA	126.96	8.39	PRO			PRO		
3	ALA	124.25	8.25	ALA	124.86	8.37	ALA	124.73	8.37	ALA	124.28	8.25	ALA	124.63	8.37	ALA	124.67	8.37
4	LYS	122.63	8.20	LYS	120.51	8.24	LYS	121.56	8.21	LYS	120.62	8.25	LYS	122.33	8.24	LYS	122.36	8.25
5	PRO			ALA	125.59	8.29	PRO			ALA	125.43	8.30	PRO			PRO		
6	GLU	121.78	8.43	GLU	120.18	8.28	GLU	120.86	8.46	GLU	119.88	8.31	GLU	121.44	8.43	GLU	121.41	8.42
7	ALA	126.91	8.37	ALA	126.03	8.18	ALA	124.91	8.22	ALA	124.98	8.24	ALA	126.85	8.27	ALA	126.81	8.27
8	PRO			PRO			ALA	122.98	8.04	ALA	123.04	8.21	PRO			PRO		
9	GLY	107.85	8.27	GLY	108.88	8.38	GLY	107.78	8.26	GLY	107.78	8.29	GLY	108.21	8.32	GLY	108.28	8.32
10	GLU	118.58	8.33	GLU	119.05	8.24	GLU	119.57	8.18	GLU	119.79	8.17	GLU	118.86	8.23	GLU	118.97	8.22
11	ASP	117.88	8.50	ASP	118.74	8.46	ASP	119.69	8.40	ASP	120.49	8.40	ASP	118.57	8.48	ASP	118.64	8.47
12	ALA	122.74	7.51	ALA	123.31	7.78	ALA	123.54	7.94	ALA	124.55	8.23	ALA	124.18	7.75	ALA	123.08	7.71
13	SER	117.69	8.39	SER	117.05	8.32	SER	116.75	8.25	SER	114.68	8.26	PRO			ALA	124.19	8.27
14	PRO			PRO			PRO			ALA	124.79	8.27	ALA	123.71	8.50	PRO		
15	GLU	117.63	8.48	GLU	117.85	8.45	GLU	118.22	8.44	GLU	118.39	8.25	GLU	117.55	8.67	GLU	118.08	8.73
16	GLU	120.86	7.89	GLU	120.66	7.92	GLU	120.56	7.97	GLU	120.09	8.05	GLU	119.47	7.84	GLU	119.75	7.92
17	LEU	121.03	8.33	LEU	120.79	8.23	LEU	121.09	8.23	LEU	120.97	8.13	LEU	121.57	7.89	LEU	121.65	7.92
18	SER	114.03	8.29	SER	113.99	8.20	SER	114.00	8.13	SER	114.05	8.09	SER	113.88	8.23	SER	114.11	8.24
19	ARG	121.43	7.91	ARG	121.36	7.87	ARG	121.68	7.89	ARG	121.72	7.97	ARG	121.46	7.92	ARG	121.65	7.98
20	TYR	121.71	8.02	TYR	120.85	8.00	TYR	120.98	8.05	TYR	121.01	8.06	TYR	121.22	8.01	TYR	121.06	7.98
21	TYR	118.82	8.53	TYR	118.85	8.36	TYR	119.16	8.31	TYR	119.20	8.31	TYR	119.07	8.36	TYR	119.20	8.31
22	ALA	121.48	8.03	ALA	121.90	8.05	ALA	121.95	8.05	ALA	121.88	8.07	ALA	121.89	8.00	ALA	122.10	8.01
23	SER	115.50	8.10	SER	114.04	7.90	SER	114.24	7.94	SER	114.05	7.93	SER	115.07	8.05	SER	114.93	8.04
24	LEU	120.14	8.38	LEU	122.42	7.64	LEU	122.78	7.77	LEU	122.60	7.70	LEU	123.67	8.14	LEU	123.64	8.11
25	ARG	118.21	7.96	ARG	118.17	7.83	ARG	118.22	7.89	ARG	118.15	7.87	ARG	118.32	7.92	ARG	118.50	7.92
26	HIS	116.69	7.93	HIS	117.12	8.04	HIS	117.23	8.03	HIS	117.25	8.05	HIS	117.03	7.97	HIS	117.16	7.99
27	TYR	120.11	8.14	TYR	119.78	8.05	TYR	119.87	8.07	TYR	119.80	8.07	TYR	120.13	8.13	ALA	120.24	8.13
28	LEU	117.89	8.01	LEU	121.54	8.14	LEU	121.45	8.18	LEU	121.67	8.15	LEU	120.71	8.32	LEU	120.99	8.31
29	ASN	117.98	7.88	ASN	118.39	8.14	ASN	118.39	8.14	ASN	118.43	8.16	ASN	120.85	8.08	ASN	120.84	8.08
30	LEU	120.84	7.77	LEU	121.55	7.92	LEU	121.50	7.91	LEU	121.59	7.94	LEU	118.17	8.06	LEU	118.27	8.08
31	VAL	121.70	8.00	VAL	119.42	7.97	VAL	119.22	7.97	VAL	119.43	7.99	VAL	121.02	7.83	VAL	121.20	7.86
32	THR	115.67	7.87	THR	116.38	7.98	THR	116.29	7.97	THR	116.31	7.99	THR	115.82	7.89	THR	116.09	7.90
33	ARG	121.96	8.02	ARG	122.31	8.12	ARG	122.26	8.10	ARG	122.28	8.12	ARG	122.07	8.04	ARG	122.20	8.06
34	GLN	120.35	8.17	GLN	120.72	8.21	GLN	120.66	8.21	GLN	120.68	8.22	GLN	120.45	8.17	GLN	120.59	8.19
35	ARG	121.23	8.17	ARG	121.60	8.21	ARG	121.55	8.23	ARG	121.57	8.22	ARG	121.35	8.18	ARG	121.48	8.19
36	TYR	120.63	8.06	TYR	120.84	8.08	TYR	120.80	8.07	TYR	120.64	8.03	TYR	120.73	8.06	TYR	120.79	8.06

Table S8: Chemical shifts for pPYY Tyr->Ala mutants, 1 mM samples in 90% H_2O/10%^2H_2O, pH 4.1, T 301K, referenced to the signal of residual HDO at 4.72 ppm.

N	Y20A-pPYY			Y21A-pPYY			Y27A-pPYY		
	Res.	N, ppm	H^N, ppm	Res.	N, ppm	H^N, ppm	Res.	N, ppm	H^N, ppm
1	TYR			TYR			TYR		
2	PRO			PRO			PRO		
3	ALA	124.77	8.37	ALA	124.60	8.38	ALA	124.77	8.37
4	LYS	121.49	8.24	LYS	122.91	8.26	LYS	121.34	8.23
5	PRO			PRO			PRO		
6	GLU	120.85	8.42	GLU	122.01	8.46	GLU	120.99	8.41
7	ALA	126.32	8.26	ALA	127.09	8.30	ALA	126.09	8.19
8	PRO			PRO			PRO		
9	GLY	109.08	8.45	GLY	107.92	8.29	GLY	108.78	8.37
10	GLU	119.61	8.13	GLU	118.70	8.31	GLU	119.17	8.22
11	ASP	120.08	8.42	ASP	118.13	8.51	ASP	118.84	8.45
12	ALA	123.86	8.05	ALA	122.75	7.53	ALA	123.35	7.78
13	SER	116.56	8.24	SER	117.45	8.37	SER	117.00	8.30
14	PRO			PRO			PRO		
15	GLU	118.59	8.41	GLU	117.76	8.46	GLU	117.97	8.45
16	GLU	120.40	8.04	GLU	120.96	7.95	GLU	120.67	7.94
17	LEU	121.17	8.13	LEU	120.87	8.35	LEU	121.09	8.24
18	SER	114.52	8.08	SER	116.05	8.31	SER	114.00	8.21
19	ARG	122.10	8.02	ARG	122.72	8.09	ARG	121.47	7.91
20	ALA	122.94	8.05	TYR	120.58	8.15	TYR	121.11	8.01
21	TYR	118.34	8.04	ALA	122.17	8.21	TYR	119.02	8.37
22	ALA	122.63	8.04	ALA	120.12	7.89	ALA	121.91	8.06
23	SER	113.40	8.02	SER	115.96	8.19	SER	114.14	7.90
24	LEU	122.84	7.91	LEU	124.11	8.41	LEU	122.55	7.58
25	ARG	118.91	8.01	ARG	118.51	7.84	ARG	118.36	7.79
26	HIS	117.72	8.14	HIS	116.94	7.99	HIS	117.50	8.09
27	TYR	120.13	8.10	TYR	120.28	8.14	ALA	123.45	8.10

28	LEU	121.94	8.18	LEU	120.17	8.43	LEU	120.41	8.12
29	ASN	118.62	8.19	ASN	118.07	8.02	ASN	118.70	8.25
30	LEU	121.69	7.96	LEU	120.70	7.78	LEU	121.78	8.01
31	VAL	119.54	8.00	VAL	117.52	7.87	VAL	120.31	8.01
32	THR	116.45	8.00	THR	115.54	7.83	THR	117.26	8.08
33	ARG	122.36	8.13	ARG	121.95	8.00	ARG	122.72	8.21
34	GLN	120.76	8.23	GLN	120.31	8.15	GLN	121.11	8.28
35	ARG	121.64	8.22	ARG	121.24	8.16	ARG	121.84	8.25
36	TYR	120.88	8.08	TYR	120.65	8.05	TYR	120.99	8.10

Table S9: Chemical shifts for pPYY Pro->Ala mutants 1 mM samples in 90% $H_2O/10\%^2H_2O$, **300mM DPC**, pH 6.0, T 310K, referenced to the signal of residual HDO at 4.63 ppm.

Res. N	P2A-pPYY			P5A-pPYY			P8A-pPYY			P2,5,8,14A-pPYY		
	Res.	N, ppm	H^N, ppm	Res.	N, ppm	H^N, ppm	Res.	N, ppm	H^N, ppm	Res.	N, ppm	H^N, ppm
1	TYR			TYR			TYR			TYR		
2	ALA	126.88	8.37	PRO			PRO			ALA	126.85	8.39
3	ALA	123.60	8.17	ALA	124.38	8.29	ALA	123.96	8.25	ALA	123.77	8.20
4	LYS	121.01	8.14	LYS	120.07	8.19	LYS	120.56	8.16	LYS	120.15	8.18
5	PRO			ALA	125.35	8.30	PRO			ALA	125.30	8.29
6	GLU	120.55	8.45	GLU	119.76	8.28	GLU	120.56	8.55	GLU	119.70	8.32
7	ALA	125.59	8.18	ALA	125.58	8.19	ALA	124.32	8.19	ALA	124.42	8.18
8	PRO			PRO			ALA	122.88	8.16	ALA	122.92	8.18
9	GLY	108.84	8.40	GLY	108.80	8.39	GLY	107.59	8.27	GLY	107.63	8.27
10	GLU	119.94	8.15	GLU	119.93	8.15	GLU	119.91	8.21	GLU	120.09	8.23
11	ASP	120.53	8.35	ASP	120.49	8.35	ASP	120.41	8.34	ASP	120.78	8.36
12	ALA	123.81	8.12	ALA	123.77	8.11	ALA	123.61	8.09	ALA	124.54	8.27
13	SER	116.38	8.31	SER	116.37	8.31	SER	116.27	8.28	SER	114.72	8.29
14	PRO			PRO			PRO			ALA	124.75	8.28
15	GLU	119.86	8.43	GLU	119.84	8.43	GLU	119.70	8.43	GLU	118.00	8.15
16	GLU	120.69	8.30	GLU	120.68	8.30	GLU	120.63	8.28	GLU	120.26	8.14
17	LEU	119.35	8.31	LEU	119.30	8.30	LEU	119.26	8.30	LEU	118.89	8.28

18	SER	113.19	8.14	SER	113.19	8.14	SER	113.16	8.14	SER	113.15	8.11
19	ARG	122.60	7.92	ARG	122.60	7.92	ARG	122.58	7.91	ARG	122.61	7.92
20	TYR	119.20	8.06	TYR	119.20	8.07	TYR	119.19	8.07	TYR	119.41	8.16
21	TYR	119.13	8.31	TYR	119.12	8.31	TYR	119.18	8.32	TYR	119.44	8.46
22	ALA	120.45	8.19	ALA	120.43	8.19	ALA	120.44	8.19	ALA	120.42	8.20
23	SER	115.20	8.05	SER	115.20	8.05	SER	115.22	8.05	SER	115.28	8.05
24	LEU	122.10	8.24	LEU	122.10	8.24	LEU	122.13	8.25	LEU	122.28	8.30
25	ARG	117.34	8.35	ARG	117.34	8.34	ARG	117.33	8.35	ARG	117.34	8.37
26	HIS	117.43	7.91	HIS	117.44	7.92	HIS	117.32	7.92	HIS	117.37	7.92
27	TYR	117.86	8.13	TYR	117.85	8.13	TYR	117.92	8.14	TYR	118.55	8.16
28	LEU	118.49	8.53	LEU	118.49	8.53	LEU	118.49	8.53	LEU	118.45	8.54
29	ASN	117.71	8.05	ASN	117.72	8.05	ASN	117.66	8.04	ASN	117.68	8.03
30	LEU	118.79	7.66	LEU	118.79	7.66	LEU	118.75	7.66	LEU	118.75	7.66
31	VAL	115.39	7.91	VAL	115.38	7.91	VAL	115.41	7.90	VAL	115.46	7.91
32	THR	110.59	7.83	THR	110.59	7.83	THR	110.60	7.83	THR	110.66	7.83
33	ARG	120.36	7.97	ARG	120.36	7.97	ARG	120.37	7.97	ARG	120.38	7.97
34	GLN	116.62	8.02	GLN	116.62	8.02	GLN	116.62	8.02	GLN	116.63	8.02
35	ARG	118.26	7.89	ARG	118.25	7.89	ARG	118.27	7.89	ARG	118.27	7.89
36	TYR	118.35	7.94	TYR	118.34	7.94	TYR	118.35	7.94	TYR	118.33	7.94

Table S10: ^{15}N{^{1}H}-NOE values for pNPY, pPYY, pPYY mutants and pPYY in MeOH 1mM samples in 90% H_2O/10%^2H_2O, pH 4.1, T 301K.

Res. N	pNPY	pPYY	P13A14-pPYY	P14A-pPYY	S13A-pPYY	P2A-pPYY	P5A-pPYY	P8A-pPYY	P2,5,8,14A-pPYY	Y20A-pPYY	Y21A-pPYY	Y27A-pPYY	A7Y-pPYY	pPYY in MeOH
2									-2.63					
3	-1.7	-0.24	-0.95	-0.43	-0.88	-1.18	-0.82	-0.55	-1.98	-1.79	-0.46	-1.45	-0.11	-2.04
4	-1.03	0.14	-0.53	-0.05	-0.99	-0.58	-0.78	-0.49	-1.61	-1.86	-0.54	-1.63	0.24	-1.15
5							-0.65		-1.30					
6	-0.8	0.41	-0.16	0.34	-0.21	0.00	-0.49	-0.50	-1.39	-0.82	0.03	-0.48	0.24	-0.91
7	-0.67	0.34	-0.21	0.19	-0.41	-0.07	-0.35	-0.35	-1.27	-1.05	-0.04	-0.85	0.43	-0.69
8								-0.38	-0.95					
9	-0.43	0.41	-0.03	0.33	-0.19	0.00	-0.14	-0.26	-0.81	-0.60	-0.01	-0.16	-	-0.44
10	-0.33	0.45	0.07	0.32	-0.39	0.09	-0.11	-0.32	-0.28	-0.70	0.03	-0.47	0.33	-0.26
11	-0.39	0.51	0.21	0.48	-0.08	0.16	-0.07	-0.20	-0.35	-0.40	0.14	0.10	0.41	-0.25
12	-0.24	0.51	0.25	0.54	0.05	0.17	-0.10	-0.04	-0.11	-0.35	0.15	-0.36	0.53	-0.26
13		0.59		0.53	-0.14	0.28	0.01	0.02	0.06	-0.47	0.30	-0.06	0.51	0.27
14	0.07		0.46	0.71					0.05					
15	0.04	0.70	0.43	0.72	0.26	0.46	0.17	0.10	0.24	0.09	0.48	0.41	0.57	0.60
16	0.15	0.70	0.47	0.69	0.32	0.48	0.16	0.19	0.17	0.08	0.51	0.35	0.66	0.46
17	0.13	0.65	0.42	0.66	0.49	0.49		0.13	0.32	-0.02	0.44	0.10	-	0.49
18	0.17	0.66	0.54	0.68	0.23	0.45	0.12	0.10	0.27	0.06	0.44	0.14	0.60	0.50
19	0.21	0.65	0.50	0.67	0.32	0.50	0.17	0.24	0.31	0.18	0.49	0.34	0.52	0.52
20	0.2	0.67	0.46	0.70	0.50	0.31	0.17	0.14	0.28	0.13	0.34	0.42	0.70	0.57
21	0.39	0.72	0.50	0.74	0.41	0.54	0.19	0.23	0.35	0.06	0.39	0.42	0.70	0.60
22	0.25	0.71	0.51	0.74	0.36	0.59	0.16	0.17	0.31	0.19	0.42	0.40	0.65	0.69
23	0.25	0.69	0.54	0.74	0.57	0.55	0.20	0.22	0.22	0.18	0.37	0.31	0.70	0.63
24	0.23	0.73	0.56	0.56	0.42	0.42	0.16	0.12	0.13	0.01	0.54	0.38	0.66	0.54
25	0.18	0.69	0.32	0.64	0.21	0.41	0.09	0.22	0.19	0.04	0.28	0.04	0.62	0.60
26	0.22	0.71	0.42	0.69	0.39	0.52	0.16	0.15	0.09	-0.03	0.53	0.15	0.64	0.56
27	0.03	0.71	0.56	0.71	0.22	0.31	0.08	0.04	-0.01	-0.03	0.28	0.05	0.64	0.62
28	-0.02	0.69	0.37	0.66	0.25	0.55	0.01	0.00	-0.04	-0.15	0.45	-0.26	0.62	0.58
29	0.1	0.59		0.55	0.30		-0.01	-0.04	-0.07	-0.24	0.48	-0.54	0.64	0.67
30	0.03	0.47	0.35	0.54	0.23	0.27	-0.03	0.05	-0.08	-0.17	0.18	-0.09	0.58	0.52
31	-0.03	0.59	0.33	0.54	0.05	0.10	-0.06	-0.06	-0.21	-0.18	0.11	-0.24	0.45	0.54
32	-0.06	0.51	0.12	0.54	-0.16	0.00	-0.12	-0.05	-0.45	-0.34	-0.05	-0.41	0.30	0.54
33	-0.21	0.27	0.14	0.29	-0.12	0.04	-0.24	-0.28	-0.90	-0.64	0.06	-1.11	0.41	0.44
34	-0.6	0.21	-0.14		-0.80	-0.41	-0.42		-1.22	-0.88	-0.22	-1.31	-0.24	0.56
35	-0.79	-0.06	-0.56	0.21	-1.23	-0.85	-0.59	-0.31	-0.92	-1.56	-0.83	-1.69	-0.68	0.35
36	-0.99	-0.55	-0.81	-0.09	-1.10	-0.95	-0.59	-0.48	-2.63	-1.36	-0.80	-1.60	-0.83	0.42
mean 17-31	0.16±0.09	0.65±0.05	0.46±0.07	0.65±0.06	0.33±0.12	0.43±0.11	0.10±0.07	0.11±0.08	0.14±0.15	0.00±0.14	0.38±0.12	0.11±0.29	0.62±0.07	0.58±0.06

Table S11: $^{15}N\{^1H\}$-NOE values for pPYY and pPYY Pro->Ala mutants 1mM samples in 90% H_2O/10%^2H_2O, **300mM DPC**, pH 6.0, T 310K.

res. N	pPYY	P2A-pPYY	P5A-pPYY	P8A-pPYY	P2,5,8,14A-pPYY
2		-			-2.31
3	-1.35	-1.68	-1.51	-1.38	-1.91
4	-1.30	-1.44	-1.45	-1.35	-1.45
5			-1.14		-1.12
6	-1.00	-1.14	-0.96	-1.25	-0.92
7	-0.92	-1.09	-1.04	-1.16	-1.21
8				-1.18	-1.13
9	-0.72	-0.76	-0.79	-0.82	-0.74
10	-0.64	-0.65	-0.69	-0.60	-0.41
11	-0.28	-0.30	-0.32	-0.33	-0.21
12	-0.41	-0.51	-0.49	-0.45	0.06
13	0.03	0.05	-0.02	-0.06	0.22
14					0.37
15	0.30	0.28	0.24	0.28	0.60
16	0.45	0.38	0.38	0.36	0.32
17	0.64	0.55	0.54	0.54	0.58
18	0.59	0.58	0.54	0.57	0.59
19	0.69	0.66	0.68	0.66	0.71
20	0.63	0.63	0.63	0.61	0.54
21	0.67	0.73	0.63	0.61	0.73
22	0.68	0.65	0.60	0.66	0.58
23	0.67	0.72	0.69	0.67	0.71
24	0.73	0.73	0.70	0.67	0.74
25	0.76	0.73	0.74	0.75	0.76
26	0.77	0.73	0.75	0.72	0.73
27	0.70	0.74	0.68	0.70	0.36
28	0.71	0.68	0.67	0.71	0.74
29	0.72	0.69	0.71	0.70	0.72
30	0.67	0.62	0.64	0.64	0.68

31	0.68	0.66	0.68	0.66	0.74
32	0.58	0.57	0.62	0.60	0.58
33	0.60	0.52	0.60	0.53	0.57
34	0.51	0.49	0.50	0.52	0.47
35	0.42	0.45	0.45	0.43	0.42
36	0.32	0.33	0.34	0.32	0.33
mean 17-31	0.69±0.05	0.67±0.06	0.66±0.06	0.66±0.06	0.66±0.11

Table S12: $^{3}J_{\mathrm{HN}\alpha}$ scalar coupling constants for pPYY, pNPY, pPYY mutants in water and pPYY in MeOH. Values ≤ 6.0 Hz are shaded in grey.

Res. N	P2A-pPYY	P5A-pPYY	P8A-pPYY	P2,5,8,14A-pPYY	P14A-pPYY	P13A14-pPYY	S13A-pPYY	pPYY	pNPY	pPYY in MeOH	A7Y-pPYY	Y20A-pPYY	Y21A-pPYY	Y27A-pPYY
1														
2	5.3 ± 0.1			7.5 ± 0.1										
3	6.9 ± 0.1	5.7 ± 0.2	5.7 ± 0.1	6.6 ± 0.1	5.9 ± 0.3	5.7 ± 0.2	5.7 ± 0.2	6.1 ± 0.2	7.3 ± 0.4	5.9 ± 0.1	5.9 ± 0.1	5.3 ± 0.4	5.9 ± 0.4	5.5 ± 0.3
4	6.9 ± 0.1	7.1 ± 0.1	8.8 ± 0.1	7.8 ± 0.1	5.3 ± 0.4	6.2 ± 0.2	6.1 ± 0.2	5.3 ± 0.3	7.7 ± 0.1	7.1 ± 0.1	5.2 ± 0.2	7.4 ± 0.2	6.3 ± 0.2	7.3 ± 0.2
5		6.2 ± 0.2		6.7 ± 0.1										
6	7.6 ± 0.1	7.5 ± 0.1	6.8 ± 0.1	8.1 ± 0.1	6.4 ± 0.4	7.1 ± 0.1	6.9 ± 0.2	7.1 ± 0.2	7.4 ± 0.2	7.7 ± 0.1	7.5 ± 0.2	7.3 ± 0.2	6.9 ± 0.2	7.1 ± 0.3
7	7.1 ± 0.1	5.8 ± 0.2	5.1 ± 0.1	6.7 ± 0.1	5.0 ± 0.4	5.5 ± 0.2	5.4 ± 0.2	5.1 ± 0.3	8.9 ± 0.6	7.6 ± 0.3	4.6 ± 0.3	6.1 ± 0.4	5.1 ± 0.4	6.1 ± 0.4
8			6.1 ± 0.2	6.2 ± 0.1										
9	7.5 ± 0.4	-	6.0 ± 0.3	6.9 ± 0.1	4.3 ± 1.4	6.5 ± 0.3	6.5 ± 0.3	7.4 ± 0.7	6.1 ± 0.7	6.1 ± 0.8	-	5.3 ± 0.9	6.2 ± 0.6	5.5 ± 0.8
10	5.0 ± 0.2	7.6 ± 0.3	6.3 ± 0.1	7.6 ± 0.1	5.1 ± 0.3	5.7 ± 0.2	5.5 ± 0.2	5.5 ± 0.3	7.0 ± 0.2	6.2 ± 0.2	4.7 ± 0.2	6.8 ± 0.3	5.2 ± 0.3	5.6 ± 0.4
11	8.9 ± 0.1	8.1 ± 0.2	8.0 ± 0.1	8.2 ± 0.1	9.5 ± 0.7	9.2 ± 0.1	8.1 ± 0.2	9.8 ± 0.1	9.0 ± 0.1	7.1 ± 0.1	-	7.4 ± 0.3	8.4 ± 0.2	8.4 ± 0.2
12	5.0 ± 0.2	5.6 ± 0.3	6.2 ± 0.2	5.2 ± 0.1	6.1 ± 0.4	5.4 ± 0.2	5.8 ± 0.3	5.0 ± 1.3	6.6 ± 0.6	6.6 ± 0.1	4.7 ± 0.4	6.7 ± 0.3	5.6 ± 0.3	5.5 ± 0.9
13	6.2 ± 0.1	6.1 ± 0.4	6.0 ± 0.1	6.2 ± 0.1	6.0 ± 0.3		5.8 ± 0.2	5.7 ± 0.4		6.4 ± 0.3	5.4 ± 0.2	6.3 ± 0.3	6.0 ± 0.4	6.0 ± 0.4
14				4.6 ± 0.1	4.5 ± 0.7	4.8 ± 0.1			5.0 ± 0.4					
15	4.9 ± 0.1	5.2 ± 0.3	4.6 ± 0.3	5.5 ± 0.1	5.1 ± 0.4	5.6 ± 0.2	5.6 ± 0.3	5.1 ± 0.3	5.2 ± 1.8	5.7 ± 0.3	4.9 ± 0.2	5.4 ± 0.3	5.6 ± 0.3	5.3 ± 0.4
16	5.9 ± 0.2	5.8 ± 0.6	5.8 ± 0.2	5.1 ± 0.1	5.9 ± 0.4	6.5 ± 0.2	6.7 ± 0.3	6.0 ± 0.5	7.1 ± 0.6	6.1 ± 0.2	6.0 ± 0.2	6.5 ± 0.3	5.8 ± 0.7	5.8 ± 0.6
17	5.4 ± 0.2	4.3 ± 0.1	5.2 ± 0.2	5.1 ± 0.1	4.2 ± 0.7	4.7 ± 0.4	5.4 ± 0.4	5.0 ± 0.4	5.7 ± 0.2	5.3 ± 0.2	5.8 ± 0.3	6.2 ± 0.4	5.0 ± 0.7	4.9 ± 0.5
18	4.1 ± 0.2	5.2 ± 0.2	4.9 ± 0.2	4.5 ± 0.1	4.3 ± 0.6	5.2 ± 0.2	4.9 ± 0.3	4.7 ± 0.5	5.2 ± 0.1	5.0 ± 0.2	4.7 ± 0.2	5.9 ± 0.3	5.5 ± 0.3	5.3 ± 0.3
19	6.8 ± 0.3	6.9 ± 0.2	5.1 ± 0.2	5.2 ± 0.1	5.1 ± 0.3	5.7 ± 0.3	5.9 ± 0.4	6.2 ± 0.2	5.2 ± 0.2	5.2 ± 0.2	-	6.0 ± 0.4	4.9 ± 0.9	5.7 ± 0.2
20	-	7.9 ± 0.1	5.4 ± 0.2	5.1 ± 0.1	6.0 ± 0.3	5.2 ± 0.2	5.4 ± 0.2	5.4 ± 0.8	6.2 ± 0.4	5.0 ± 0.2	4.4 ± 0.3	6.0 ± 0.3	4.3 ± 1.3	5.0 ± 0.5
21	4.6 ± 0.4	5.6 ± 0.3	5.5 ± 0.3	5.3 ± 0.2	4.8 ± 0.4	5.1 ± 0.2	5.2 ± 0.3	5.3 ± 1.0	6.6 ± 0.3	4.8 ± 0.2	4.8 ± 0.3	-	5.0 ± 0.4	5.5 ± 0.5
22	5.6 ± 0.4	5.9 ± 0.1	4.8 ± 0.1	4.8 ± 0.1	5.2 ± 0.1	5.0 ± 0.2	5.4 ± 0.2	4.9 ± 0.3	5.8 ± 0.4	4.9 ± 0.2	5.8 ± 0.3	5.8 ± 0.2	5.5 ± 0.3	5.4 ± 0.1
23	5.6 ± 0.1	8.3 ± 0.2	6.0 ± 0.2	5.7 ± 0.1	5.2 ± 0.5	5.7 ± 0.2	5.9 ± 0.4	6.0 ± 1.4	5.8 ± 0.5	5.1 ± 0.2	5.1 ± 0.2	6.1 ± 0.4	4.9 ± 0.9	6.0 ± 0.4
24	5.1 ± 0.1	7.1 ± 0.3	5.3 ± 0.3	6.5 ± 0.4	6.6 ± 0.4	6.0 ± 0.3	6.5 ± 0.3	6.0 ± 0.2	6.3 ± 0.5	5.0 ± 0.2	4.8 ± 0.5	6.3 ± 0.6	6.4 ± 0.4	6.1 ± 0.4
25	6.7 ± 0.1	6.1 ± 0.2	5.7 ± 0.3	6.0 ± 0.2	4.9 ± 0.5	7.1 ± 0.2	6.7 ± 0.3	5.3 ± 0.9	5.4 ± 1.5	5.0 ± 0.2	4.9 ± 0.2	6.6 ± 0.4	5.3 ± 0.4	6.5 ± 0.2
26	5.6 ± 0.3	7.1 ± 0.3	6.3 ± 0.2	6.9 ± 0.1	5.4 ± 0.6	6.0 ± 0.2	6.0 ± 0.4	5.6 ± 0.2	6.5 ± 0.4	4.9 ± 0.2	4.8 ± 0.5	6.6 ± 0.4	5.5 ± 0.4	7.0 ± 0.4
27	7.0 ± 0.1	7.8 ± 0.3	6.0 ± 0.2	6.5 ± 0.1	5.5 ± 0.2	5.8 ± 0.2	6.2 ± 0.3	6.1 ± 0.4	6.5 ± 0.3	4.9 ± 0.2	4.9 ± 0.3	6.3 ± 0.5	-	5.7 ± 0.4
28	6.9 ± 0.5	5.3 ± 0.5	5.7 ± 0.2	6.4 ± 0.1	5.9 ± 0.3	5.6 ± 0.2	5.9 ± 0.3	4.4 ± 1.2	7.1 ± 0.2	4.7 ± 0.2	5.3 ± 0.2	6.7 ± 0.5	5.5 ± 0.5	6.3 ± 0.4
29	5.0 ± 0.2	6.9 ± 0.5	6.8 ± 0.2	7.1 ± 0.1	5.9 ± 0.3	5.0 ± 0.2	-	5.6 ± 0.4	5.9 ± 0.3	5.0 ± 0.2	5.4 ± 0.2	6.8 ± 0.5	5.8 ± 0.8	7.6 ± 0.2
30	6.7 ± 0.1	5.7 ± 0.6	6.8 ± 0.2	6.7 ± 0.2	6.3 ± 0.3	6.3 ± 0.2	6.1 ± 0.3	6.3 ± 0.3	6.5 ± 0.4	5.1 ± 0.2	6.3 ± 0.2	6.9 ± 0.8	6.2 ± 0.5	6.8 ± 0.3
31	7.3 ± 0.1	7.0 ± 0.2	8.7 ± 0.4	7.7 ± 0.1	6.6 ± 0.3	6.4 ± 0.2	6.6 ± 0.3	6.3 ± 0.1	6.6 ± 0.6	5.0 ± 0.2	6.5 ± 0.2	7.3 ± 0.4	6.9 ± 0.3	7.9 ± 0.2
32	6.4 ± 0.1	7.5 ± 0.2	8.6 ± 0.1	7.7 ± 0.1	7.1 ± 0.3	6.3 ± 0.1	7.0 ± 0.3	6.8 ± 0.3	6.9 ± 0.3	5.1 ± 0.2	6.5 ± 0.2	7.2 ± 0.4	6.2 ± 0.3	8.0 ± 0.1
33	6.9 ± 0.1	6.8 ± 0.3	6.2 ± 0.1	7.2 ± 0.1	6.3 ± 0.2	7.2 ± 0.3	6.7 ± 0.3	6.7 ± 0.3	6.9 ± 0.3	5.2 ± 0.2	-	6.6 ± 0.6	6.7 ± 0.4	7.0 ± 0.3
34	-	6.6 ± 0.3	6.2 ± 0.1	7.6 ± 0.1	6.6 ± 0.2	6.4 ± 0.2	6.4 ± 0.2	6.6 ± 0.2	6.0 ± 0.6	5.3 ± 0.2	6.2 ± 0.1	6.7 ± 0.3	7.1 ± 0.6	6.7 ± 0.1
35	7.3 ± 0.1	6.9 ± 0.1	6.7 ± 0.1	7.3 ± 0.1	6.7 ± 0.2	6.9 ± 0.1	6.8 ± 0.2	6.8 ± 0.2	7.5 ± 0.2	6.0 ± 0.2	6.6 ± 0.1	7.4 ± 0.2	6.8 ± 0.3	7.1 ± 0.2
36	8.6 ± 0.1	8.9 ± 0.1	9.0 ± 0.1	8.6 ± 0.1	8.6 ± 0.1	8.8 ± 0.1	8.9 ± 0.1	8.9 ± 0.2	9.3 ± 0.1	9.3 ± 0.1	8.4 ± 0.1	8.6 ± 0.2	8.6 ± 0.2	9.1 ± 0.1

References:

[1] Blundell, T. L., Pitts, J. E., Tickle, S. P. & Wu, C. W. (1981) X-ray analysis (1. 4-Å resolution) of avian pancreatic polypeptide: Small globular protein hormone, *Proc. Natl. Acad. Sci. U S A* 78, 4175-9.

[2] Lerch, M., Mayrhofer, M. & Zerbe, O. (2004) Structural similarities of micelle-bound peptide YY (PYY) and neuropeptide Y (NPY) are related to their affinity profiles at the Y receptors, *J. Mol. Biol.* 339, 1153-68.

[3] Becke, A. D. (1993) Density-functional thermochemistry III. The role of exact exchange, *J. Chem. Phys.* 98, 5648-52.

[4] Schrödinger, Inc., Portland, Oregon, Jaguar 4.2, 2000.

6. Studies of unfolding of PYY in water and methanol by NMR

Unfolding of the 36 amino acid polypeptide PYY was studied by heat- and solvent-induced denaturation. Structural changes were followed using ^{13}C-NMR spectroscopy. Moreover, the structure of PYY in water was re-determined to high resolution using three-dimension ^{13}C-resolved NOESY data as well as a larger set of residual dipolar couplings (RDCs): Moreover, the peptide in methanol solution is characterized both by determining its solution structure as well as by its internal backbone dynamics as derived from ^{15}N relaxation data. The latter is characterized by a complete loss of tertiary structure. The C-alpha chemical shifts in the heat-denaturation experiments display sigmoidal curves with very similar points of inflection indicating that both secondary as well as tertiary structure in the heat denaturation is lost synchronously. In addition, chemical shifts of selected sidechain resonances follow very similar curves, and the disruption of the structure of PYY is triggered by both larger changes in backbone as well as sidechain dihedrals. The data clearly reveal cooperative folding behavior of PYY at single-residue resolution and classify the peptides as a two state-folder. In contrast, the curves during solvent-induced denaturation display a more complicated behavior, which most likely reflects two trends that partially compensate each other: At low low-to-medium contents of methanol tertiary structure is disrupted leading to a loss of stability of secondary structure as well. At high content of methanol secondary structure is again rigidified, as clearly seen in the structure of PYY in pure methanol, which is supported by data on internal backbone dynamics in that solvent.

manuscript in preparation: Mareš, J., Walser, R., Neumoin, A., Zou, C., Zerbe, O.

Introduction

Folding of peptides and proteins has been an intense matter of research in the last decade [1]. Prediction of the structure from the amino acid sequence is still very difficult, and in the absence of high homology prone to large errors. Therefore, a much better understanding of molecular determinants of particular folds would be highly desirable. Recently, we have proposed that the peptide YY (PYY) presents a convenient system to study basic aspects of protein folding, and for formation of helical hairpins in particular [2]. The particular fold of PYY was observed by Blundell for avian PP for the first time by crystallography [3]. It is characterized by a C-terminal α-helix, which is connected via a turn to a N-terminal polyproline type-II helix. A hydrophobic contact between the α-helix and the polyproline helix leads to formation of the hairpin-type structure. Using an extensive set of single points mutants we could recently map the determinants of the PP fold. In particular, we have investigated the role of individual Pro and Tyr residues that are involved in forming the hairpin interface. Our data revealed that the presence of each of these Tyr or Pro residues is crucial, and that replacement of any of them resulted in significant destabilization of the helical hairpin. In addition, mutation of Ser13 in the turn region resulted in complete loss of tertiary structure, additionally emphasizing the role of the amino acid sequence in the turn region.

Recently, much interest has been focused on systems that do not fold cooperatively, and for which different curves are observed for the chemical shifts vs. temperature, e.g. for proteins with small or no barriers for folding that follow so-called downhill folding pathways. The group of Munoz has followed backbone and side-chain proton chemical shifts for the 40-residue polypeptide BBL, which revealed different transition temperatures at different positions along the chain, allowing them to conclude that it belongs to the class of downhill folding proteins [4]. To verify that the similarly sized PYY in contrast is really folding in cooperative (two-state) manner we have sought to record a more extensive set of probes. In this work we specifically address this question by investigating the temperature and solvent-induced denaturation of PYY in detail. Accordingly we have expressed PYY in ^{13}C-labelled form and looked at the temperature profiles of ^{13}C chemical shifts for each residue. We primarily used changes in ^{13}C-alpha chemical shifts that are almost exclusively dependent on ϕ/ψ dihedral angles to characterized changes in backbone conformational preferences.

During the mutagenesis work we have also detected a non-anticipated effect of replacing Ser-13, a residue from the turn region, by Ala [2]. Unfortunately, the structural reason for this effect was unclear, mainly because the solution structure of PYY was ill-defined in the turn region due to a restraint sampling problem. To improve the precision of the structure we refined it using ^{13}C-resolved NOESY data in combination with a set of residual dipolar couplings. Moreover, the structure of PYY in methanol is also determined, and its structural and dynamical properties compared to the peptide in water. These two structures define the initial and final states of the solvent-induced denaturation process. Changes of structure-dependent parameter such as chemical shifts in combination with these structures allowed deriving a comprehensive picture of unfolding. In addition, it firmly pinpoints cooperativity in folding of PYY in water, and suggests that tertiary structure of PYY can be selectively removed in methanol while only slightly altering secondary structure. Interestingly, for many residues the ^{13}C chemical shift profiles for the solvent denaturation curves are not monomodal, implying more complicated trajectories during the unfolding process. We believe that the data nicely demonstrate cooperativity of folding in water at a resolution that is not accessible using other spectroscopic techniques.

6.1 Results

A refined structure of PYY in water from 13C NOESY data and RDCs

Structures of porcine PYY (pPYY) in solution have been determined by Keire et al [5] and by Lerch et al [6]. Both structures have confirmed that pPYY assumes a helical hairpin. However, the structure of Keire possessed an unusual kink in the helix around residues, which was absent in the structure from our laboratory. In addition, Poulsen et al. have solved the solution structure of human PYY [7], and found it to be highly similar to the structure determined by us previously. While the helix in our previous structure was well defined, the fold of the polypeptide sequence in the turn region was much less clear and characterized by higher values of the RMSD. Nevertheless, values of the ^{15}N{^{1}H}-NOE indicated that the turn region is comparably rigid, and indicated that the low resolution is primarily a sampling problem related to the low number of medium-range NOEs in that region. In this work we have now recalculated the solution structure of PYY, using an extended set of upper-distance restraints derived from the additional usage of 3D ^{13}C-NOESY data. The 3D ^{13}C-resolved NOESY in addition to the 2D NOESY resulted in a total of 366 meaningful upper distance constraints compared to 261 in the previously reported structure calculation. Furthermore, additional restraints were derived from measuring residual dipolar couplings in stretched polyacrylamide gels involving 24 ^{15}N,^{1}H RDCs, 24 C'-CO RDCs and 10 Cα-Hα RDC (peak overlap and problems with the spectral quality in the Cα-Hα region limited the number of useful correlations).

A comparison of the refined structure of PYY in three different environments, in aqueous solution, in methanol and when bound to dodecylphosphocholine micelles, is depicted in Fig.1. While the RMSDs for backbone and all heavy atoms in the helix region encompassing residues 14-31 after superimposing backbone atoms of the same residues were 0.41±0.15Å (0.78±0.30Å) and 1.53±0.26 Å (1.69±0.39Å), respectively, and hence were slightly but not dramatically improved, the precision in the turn region (residues 5-14) after fitting backbone atoms of residues 5-31 was significantly increased in the refined dataset with values of 0.93±0.56 Å (vs. 1.77±0.64 Å) and 1.97±0.81 Å (vs. 2.74±0.94 Å) (in brackets the RMSD values in the older data set 1RU5 are shown for comparison).

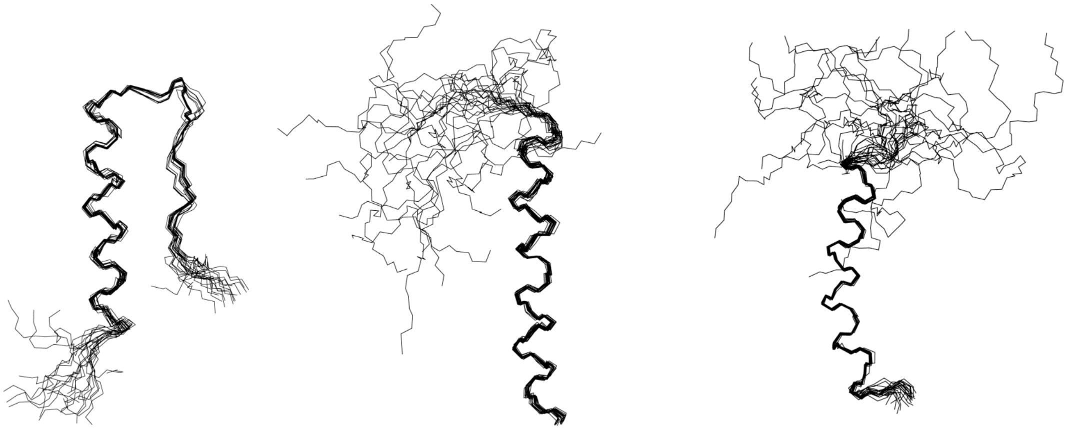


Figure 1: Backbone representation of the superposition of 20 low-energy conformers of PYY in water (left), methanol (middle) and when bound to DPC micelles (right). Backbone atoms of residues 15 to 31 have been superimposed for best fit, for PYY in water residues 4 to 8 have been included in addition.

The much-improved resolution in the β-turn region now allows comparing the solution structure of PYY with the crystal structure of aPP as determined by Blundell in 1981 [3]. He and his colleagues have postulated the presence of a β I turn in the segment Gly9 to Ala12. Unfortunately, even in the new dataset the dihedrals in that region display significant spread in their values, but average values for phi (and psi) dihedrals of residues 10 and 11 are -62° (-4°) and -85° (23.4°), respectively, and are closer to the ideal values of -60°(-30°) and -90°(0°) of a β 1 turn than to any other β turn. We like to emphasize here that the $^{15}N\{^{1}H\}$-NOE is considerably lower in the turn region (about 0.5) than in the helical region (about 0.7) so that increased flexibility may to some extent account for the larger spread in computed dihedrals.

Recently we have seen that replacement of Ser13 by Ala resulted in complete removal of the helical hairpin[2]. *Ab-initio* calculations indicated that the hydroxyl oxygen from the sidechain might be involved in a hydrogen bond formed with backbone and sidechain atoms of Glu13. While a few conformers of the energy minimized NMR ensemble did display a H-bond with the backbone amide of Glu16 no such contacts with the sidechain of Glu16 are seen.

The conformation of PYY in methanol is characterized by complete loss of tertiary structure

Unfolding the hairpin in PYY will expose hydrophobic surfaces to solvent, and the free energy of the unfolded state should hence critically depend on the hydrophobicity of the solvent. We have recently reasoned that tertiary contacts in PYY should be destabilized in the presence of a more hydrophobic solvent. We have therefore investigated whether the non back-folded state of PYY could be stabilized when the peptide is dissolved in a solvent of increased lipophilicity such as methanol. Alcohols have been utilized as a solvent for structural studies of amphipatic peptides such as neuropeptide [8]. It has also been realized that co-addition of alcohols stabilizes helical secondary structures while disrupting weak hydrophobic contacts thereby destroying tertiary structures. The order of effectiveness for destabilizing tertiary structure is trifluoroethanol > propanol > ethanol > methanol [9, 10]. The less dramatic changes observed with methanol indicate that this solvent may be particular useful to monitor the transition [11]. Mixtures of water and methanol with methanol contents of about 30% have often been found to result in molten-globule-type structures that possess a considerable extent of native secondary structure, but in which the tertiary structure has been largely destroyed [11-19]. The group of Daggett has used MD calculation in 60% methanol to simulate unfolding of ubiquitin under these conditions [20].

The structure of pPYY in 100% d_3-methanol as calculated from upper-distance limits derived from NOEs using standard 2D NMR methods is depicted in Fig 1, and compared to the structure of PYY in water or PYY when bound to phospholipids micelles. Obviously, the helical hairpin is completely removed. Not a single NOE supporting an association of the N terminus with the C-terminal helix was observed in the spectra. Nevertheless, the α-helix is stably formed and extends from residues 14 to 33, using the criteria of Kabsch and Sanders [21] for identification of secondary structure, and residues 33-36 also occupy helical backbone dihedrals. This is in contrast to the peptide in aqueous solution, where residues 32 to 36 are much more flexible.

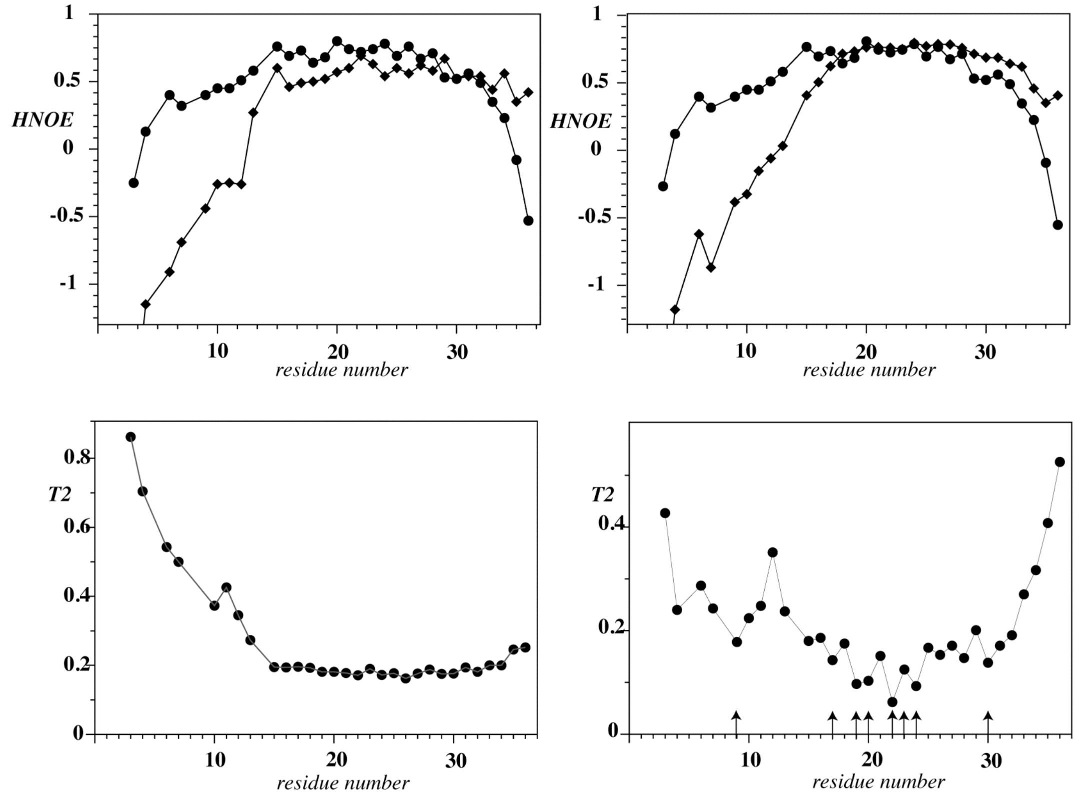


Figure 2: Comparison of $^{15}\{^{1}H\}$-NOE (top) and transverse relaxation times (T2, bottom) of PYY in various environments. Top: Data in water (circles) and in methanol (squares) (left) and in water (circles) and when bound to DPC micelles (squares) (right). Bottom: Transverse relaxation times (*T2*) of amide ^{15}N nuclei at 500 MHz, 301K, 1mM for PYY in methanol (left) and in water (right).

The values of the $^{15}N\{^{1}H\}$-NOE of PYY in pure methanol are depicted in Fig. 2 and support the view that PYY in methanol is not back-folded. The differences in internal backbone dynamics of PYY in water relative to its micelle-bound state DPC micelles compared to PYY in water vs. PYY in methanol are small indicating that the molecule undergoes similar structural transitions in both cases. The NOE values of PYY in methanol for residues preceding Ser13 are negative and increasing (negative) values towards the N terminus indicate a freely diffusing N-terminal segment. This behavior is also reflected in the *R2* rates for transverse ^{15}N relaxation of the amide ^{15}N nuclei (see Fig. 2): Interestingly, the *R2* values for residues 14 to 31 comprising the core part of the C-terminal α-helix are very uniform in PYY dissolved in methanol. In contrast, *R2* for the same residues of PYY dissolved in aqueous acetate buffer display considerable variability. We have attributed those fluctuations to exchange contributions to *R2* stemming from motions of the N-terminal segments, and this view is supported by the fact that additional exchange terms R_{ex} were required to successfully model the order parameter S^2 in the modelfree analysis for amide moieties of aqueous PYY in that segment. Our *R2* data also seem to indicate that PYY in methanol displays little tendency to self-associate. The overall correlation time was estimated from the *R2*/*R1* ratio of residues within the helix, and resulted in a value for τ_c of 3.90 ± 0.20 ns, slightly less than the 4.52 ± 0.04 ns for PYY in aqueous solution, and the difference may be due to the reduced viscosity of methanol.

Studies of Unfolding

Thermal Unfolding:

The helical hairpin in PYY is stabilized by a number of weak contacts, mainly involving Tyr residues from the C-terminal α-helix and Pro residues from the N-terminal part. We could recently demonstrate that single point mutations of these residues to Ala result in significant destabilization of the tertiary structure of PYY, indicating that formation of tertiary structure in PYY occurs in a cooperative manner. In this work we now look at thermal and unfolding of PYY in a detailed manner.

The classical experiment for folding studies is temperature-induced unfolding/folding and hence we have monitored formation of the helical hairpin by ^{13}C-NMR. The ^{13}C-alpha chemical shifts are conformation-dependent, and their deviations from random coil chemical shifts have been used to identify elements of regular secondary structure, or to improve convergence of structure calculations in the early phases of NOE assignments [22-25]. In contrast to other techniques such as fluorescence-based that monitor the overall behavior of a protein the temperature dependence of the ^{13}C-alpha chemical shifts present a convenient probe *with residue resolution* and hence directly probe for cooperativity. Temperature profiles of $^{13}C\alpha$ chemical shifts have recently been measured by Sadqui et al [4] and used to demonstrate downhill folding for the protein BBL from *E. Coli*, although this work has stirred an intense controversy [26, 27].

The temperature profiles of Cα chemical shifts for all residues extracted from constant-time [^{13}C,^{1}H]-HSQC experiments, are depicted in Fig. 3 for all those residues for which they could be extracted over the full temperature range with good fidelity. All curves are sigmoidal, and even more importantly, mono-modal, supporting the view that the backbone of PYY unfolds in highly cooperative manner. In all cases the midpoint of transition is about 315K, indicating a simultaneous loss of structure throughout the whole polypeptide. No indications are present in the data that certain parts of the sequence undergo conformational transitions that are uncoupled from the remainder.

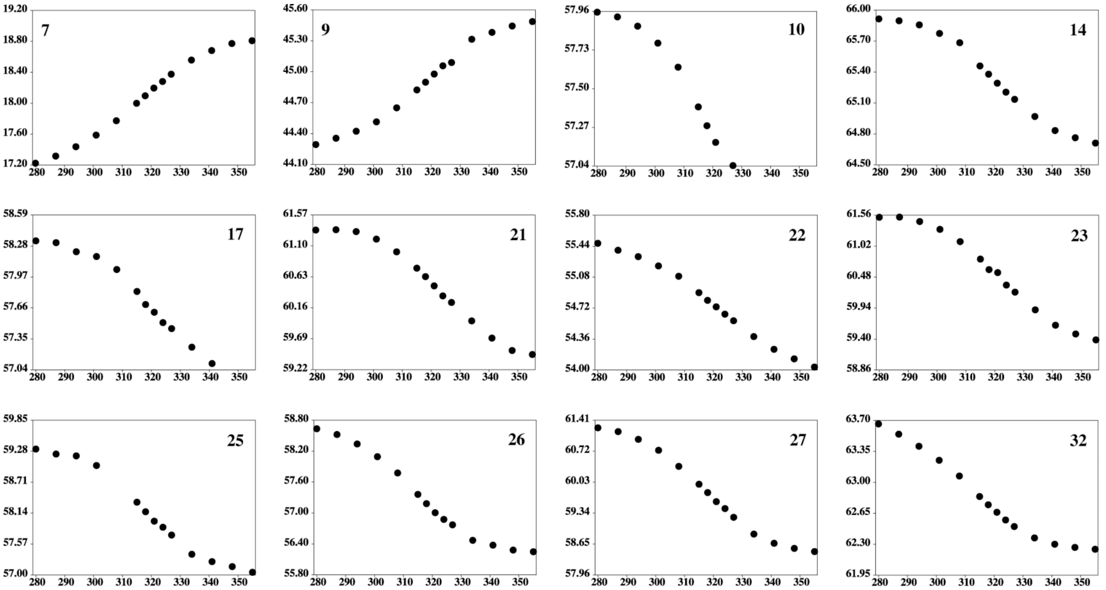


Figure 3: Cα chemical shifts of selected residues of PYY in water at various temperatures with residue numbers annotated in the figures.

Tertiary structure in PYY in water is stabilized by interactions between Tyr residues 20 and 27, located on the hydrophobic side of the C-terminal helix, and Pro residues 2, 5 and 8 in the N-terminal polyproline segment. Loss of tertiary structure may therefore also influence the sidechain conformations of these Tyr residues. Unfortunately, most of the C-beta resonances are highly overlapped, and their chemical shifts cannot be followed. For the isolated C-β resonance of Tyr21 a profile very similar to the profiles from Ca of Tyr21 is observed. (see Fig. 4), and a very similar behavior is also seen for the proton chemical shift difference of the two β-protons of Tyr21. We conclude that the sidechain behavior of Tyr21, a residue that is of prime importance for formation of the helical hairpin, changes its conformational preferences synchronously with the backbone of the polypeptide. This observation is in accordance with changes in proton chemical shifts of the methyl group of Ala-7 and Hδ of Tyr20, presented by us earlier [2].

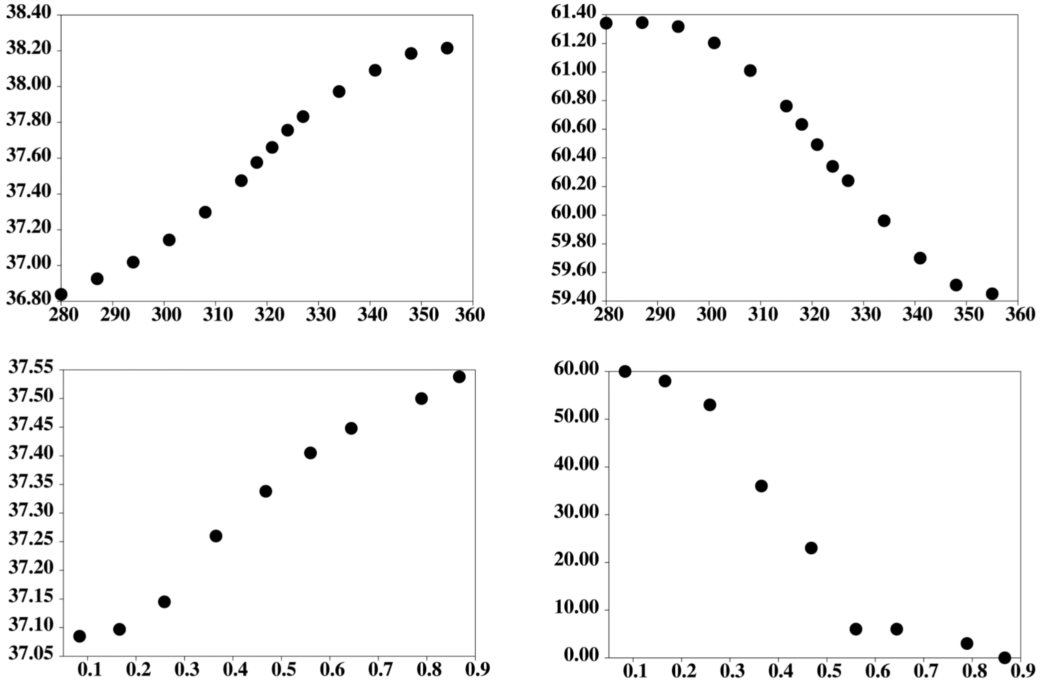


Figure 4: Side-chain chemical shift changes of PYY in heat-induced (top) or solvent-induced (bottom) denaturation experiments. Left: C-β chemical shifts of Tyr21 Right: Difference of proton chemical shifts of the two Hβ-resonances of Tyr21.

Solvent-induced Unfolding:

Removal of contacts between N- and C-terminal residues that establish the helical hairpin is expected to result in complete loss of structure for all residues preceding the first helical residue Pro14. An interesting observation therefore is that the segment comprising residues 12 to 14 apparently is somehow better defined than expected. The N-terminal segment, although largely unstructured, does not fully sample conformational space, and its presence is limited to the space defined by vicinity to the helix and never occurs in the region extending away from the helix. In our previous mutational study we observed that replacement of Ser13 by Ala resulted in loss of tertiary structure, and speculated that the hydroxyl group might be involved in formation of a hydrogen bond that helps stabilizing the conformation of the turn region. In fact, we proposed that the turn, although stabilized by the tertiary contact, is present to some extend independent of the latter. Nevertheless, scalar coupling constants for residues of the turn region are all well-above 6 Hz (see Supplementary Material) indicating that any structure in that region is transient.

Similar to the thermal unfolding studies we have probed the solvent-induced structural transition when the solvent is gradually changed from pure water to pure methanol by monitoring the ^{13}C-alpha chemical shifts. It should be noticed, however, that unfolding in methanol is principally different in nature when compared to thermal unfolding in water: Unfolding in water leads to the complete loss of structure (*vide supra*) while unfolding in methanol results in removal of tertiary structure while secondary structure persists or is even stabilized.

Representative profiles of ^{13}C-alpha chemical shift vs. solvent composition are displayed in Fig. 5. The largest changes in chemical shift ($\Delta\delta > 1$ ppm) are observed in the N-terminal region, particular in the turn region encompassing residues 8 to 13, and in the segment between residues 30 and 33. Interestingly, residues that are part of the turn region display sigmoidal curves indicating that the loss of structure in that region occurs cooperatively. Many other profiles, however, are bimodal and we observe curves with at least two points of inflection indicating that structural changes occur according to a more complex trajectory. This is particularly evident for the profiles of residues Glu16, close to the N-terminal end of the helix, and residues Tyr21 and Ala22, but is also seen for many more residues. We like to emphasize here that in contrast to the thermal denaturation experiment, which lead to the overall loss of both secondary and tertiary structure, the methanol-induced denaturation is more complicated. Firstly, tertiary structure is lost, an event that is nicely observed in the solvent profiles of residues 10 and 14. The removal of tertiary structure is expected to also result in significant destabilization of the C-terminal helix, as was demonstrated by us using single-point mutants of PYY previously [2]. The structure of PYY in pure methanol, however, has demonstrated that secondary structure in that environment is stabilized at the termini of the helix, and the C-terminal pentapeptide, that is largely flexible in aqueous solution, becomes helically structured. These two counter-acting effects are likely to occur simultaneously and are expected to result in complicated profiles for the Cα chemical shifts *vs*. solvent composition.

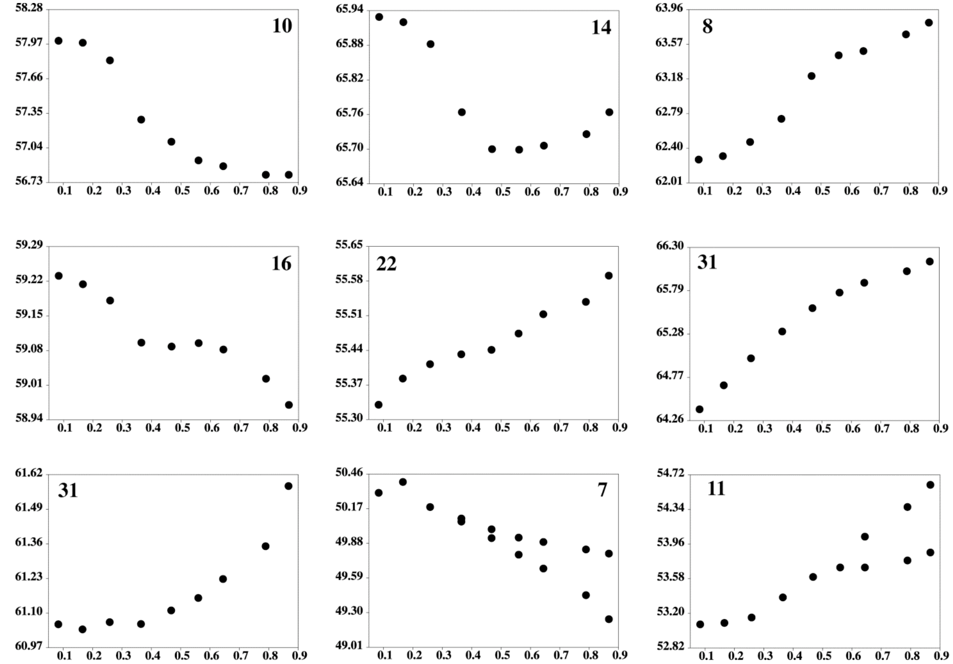


Figure 5: Cα chemical shifts of selected residues of PYY at various water/methanol mixtures (w/w) (1mM, 500 MHz ^{1}H, 301K).

Again, we have followed the isolated C-β resonance of Tyr21 in the various solvent mixtures and obtained a profile very similar to the bimodal curves observed for some of the Cα resonances is seen (see Fig. 4). In addition, the chemical shift difference of the two β-protons of Tyr21 follows a characteristic curve with an inflection point around 35% methanol.

6.2 Discussion

We have recently proposed that the peptide PYY may serve as a mini-protein, and that it posses many properties of globular proteins. The structure of the peptide is surprisingly well defined for such a monomeric short peptide that lacks disulfide bonds or other restraints. Considering its small size data can be recorded at high precision, and observables can be more easily deconvoluted into the underlying effects. Thereby, it becomes a very useful model to study some of the basic concepts of protein folding.

The mutagenesis study performed by us previously indicated that PYY displays high cooperativity in its folding behavior. Using $^{13}C\alpha$ chemical shifts we have shown here that tertiary and secondary structure is destroyed simultaneously in thermal denaturation, whereas secondary structure in methanol-induced denaturation is retained. The dissimilar behavior is due to the differences in the underlying processes of solvent- and thermal-denaturation: In the latter experiments all non-covalent contacts are destroyed, whereas in methanol-induced unfolding experiments only weak hydrophobic contacts are attenuated such that secondary structure is often retained. The comparison of thermal and solvent-induced denaturation experiments therefore allows addressing the importance of hydrophobic contacts for tertiary structure formation experimentally.

The mutagenesis experiments revealed a large role of the turn conformation for adopting the PP fold in PYY, and we proposed that the turn serves as a folding nucleus during hairpin formation. In particular, replacement of Ser13 in PYY by Ala resulted in complete loss of tertiary structure. We speculated that the sidechain of Ser13 may form a hydrogen bond with either backbone or sidechain atoms of Glu15. The now available refined structure of PYY is compatible with an H-bond to the backbone-atom of Glu15, but not to the sidechain amide moiety. The solvent-denaturation profiles of residues 10 and 14 display sigmoidal curves, indicating that structural changes in the turn region occur cooperatively. Apart from the importance of the turn region the solvent unfolding studies also support the importance of multiple weak contacts formed between Tyr 21 and 27 with the N-terminal Pro residues 2,5 and 8. PYY(3-36), a truncated version of PYY, that lacks the Pro2-Tyr27 contact, still forms weak contacts between the N-terminal part and the C-terminal helix [7], but the hairpin is significantly destabilized such that tertiary structure was observed only through a very few long-range NOEs. Changes in solvent composition leading to altered lipophilicity of the solvent reduces the magnitude of the favorable Tyr-Pro contacts, and leads to similar changes compared to those arising from the single point Pro-Ala or Tyr-Ala mutations.

Previously, we proposed that some of the destabilized mutants of PYY may serve as models for the molten globular state of PYY. Molten globular forms of polypeptides are characterized by i) structural ensembles that posses overall similarity with the folded state and ii) much more poorly packed sidechains. Mutants of enzymes behaving as molten globules may display almost wild-type biological activity [28]. Our present data suggest that PYY at certain water-to-methanol mixtures in the 30-40% range may actually also behave like a molten globule. In such an environment it becomes amenable to a more detailed structural study without the requirement for changes in the primary structure of the peptide.

The comparison of the structure of PYY in methanol to that in aqueous solutions of DPC micelles reveals striking similarity in secondary structure. In both environments no tertiary structure is observed, and the C-terminal α-helix extends over the same region. The most apparent difference in conformation is a bent of the helix in the presence of the DPC micelles, which we have attributed to the curved micelle surface. Otherwise, methanol seems to be a good mimic of the environment in a water-membrane interfacial region, and we have proposed the usage of water-methanol mixtures to mimic the transition when the peptide traverses from bulk solution into the interface. It was also realized that the properties near the water-membrane interface are significantly changed with respect to bulk solution and that such an environment may result in partial unfolding [12, 15] similar to the one observed in these water-alcohol mixtures. The membrane-potential leads to a lowered apparent pH in vicinity of the membrane surface. More importantly, there is a sharp decrease in the average dielectric constant. Since denaturing of the proteins seems to depend on average dielectric properties rather than on specific interactions we felt justified to mimic the transition from bulk solution into the water-membrane interface with water-methanol mixtures of varying relative contents.

6.3 Materials and Methods

Materials

$^{15}NH_4Cl$ and ^{13}C-glucose was purchased from Spectra Stable Isotopes (Columbia, USA) and deuterated solvents were received from Cambridge Isotope Laboratories (Andover, Massachusets, USA). The α-amidating monooxygenase used for the enzymatic conversion of Gly was purchased from Unigene. The KSI fusion plasmid was derived from the pET31b(+) vector (Invitrogen). ^{13}C-labelled PYY was expressed as described in more details previously [2, 29]. After purification by RP-HPLC it was characterized by MS: theoretical mass 4484.6 Da; experimental mass 4483.9 Da.

NMR spectroscopy and structure determination

The structure elucidation of pPYY in solution was performed on an approximately 1mM sample uniformly enriched in ^{15}N and ^{13}C at 28°C, pH 4.2 (uncorrected meter reading) in 20mM deuterated acetate buffer both in 90% H_2O/10% ^2H_2O and 99% ^2H_2O. ^{15}N relaxation data were recorded on 1mM uniformly ^{15}N labeled samples at similar conditions of pH and temperature. Structural data were measured on a Bruker AV-700 and relaxation data were recorded on a Bruker DRX-500. The $^{15}N\{1H\}$-NOEs were recorded at 50.68 MHz. Procedures for spectroscopy, data processing and data evaluation were identical to those described previously. Residual dipolar couplings were measured in stretched polyacrylamide gels using a standard IPAP (^{1}H,^{15}N) sequence [30] or J-modulation experiments [31] for measuring ^{13}Cα,Hα or C',Cα couplings. In all cases peak heights were derived from fitted Gaussians in SPARKY [32] and fitted to

$$I(\delta)=\cos(2*\pi*J(C,H)*(T-\delta))$$

in which T denotes the length of the constant-time period (14ms).

Resonances were assigned following the sequence-specific sequential resonance assignment methodology developed by Wüthrich and coworkers [33]. ^{13}C resonances were assigned from the known proton assignments and transferred with the help of [^{13}C,^{1}H]-ct-HSQC [34] spectra and (H)CCH [35] experiments. Structures were calculated using upper distance and torsion angle restraints derived from a 120ms 3D ^{13}C-NOESY spectra using the program CYANA [36]. After energy minimization using the AMBER94 all-atom force field [37] with explicit water in the program AMBER6 the 20 lowest energy structures were superimposed over backbone atoms of residues 5-7 and 17-31 and deposited in the Research Collaboratory for Structural Bioinformatics protein database under accession code 2OOO. The conformers were analyzed, including calculation of RMSD values, and figures were prepared within the program MOLMOL [38].

The data for the ^{15}N{1H}-NOE and R2 of PYY in water and methanol were recorded on 1mM solutions of the peptides using a proton-detected version of the ^{15}N{1H} steady-state NOE experiment [39] and the Carr-Purcell-Meiboom-Gill [40], respectively. The relaxation rate constants *R2* were determined by fitting volume integrals from spectra recorded for various settings of relaxation delays to three parameter single-exponential functions using the Levenberg-Marquardt algorithm [41].

References:

1. Fersht, A. R. & Daggett, V. (2002) Protein folding and unfolding at atomic resolution, *Cell. 108*, 573-82.
2. Neumoin, A., Mares, J., Lerch-Bader, M., Bader, R. & Zerbe, O. (2007) Probing the Formation of Stable Tertiary Structure in a Model Miniprotein at Atomic Resolution: Determinants of Stability of a Helical Hairpin, *Journal of the American Chemical Society*, .
3. Blundell, T. L., Pitts, J. E., Tickle, I. J., Wood, S. P. & Wu, C. W. (1981) X-ray analysis (1. 4-A resolution) of avian pancreatic polypeptide: Small globular protein hormone, *Proc. Natl. Acad. Sci. U S A. 78*, 4175-4179.
4. Sadqi, M., Fushman, D. & Muñoz, V. (2006) Atom-by-atom analysis of global downhill protein folding, *Nature. 442*, 317-21.
5. Keire, D. A., Kobayashi, M., Solomon, T. E. & Reeve, J. R. (2000) Solution structure of monomeric peptide YY supports the functional significance of the PP-fold, *Biochemistry. 39*, 9935-42.
6. Lerch, M., Mayrhofer, M. & Zerbe, O. (2004) Structural similarities of micelle-bound peptide YY (PYY) and neuropeptide Y (NPY) are related to their affinity profiles at the Y receptors, *J Mol Biol. 339*, 1153-68.
7. Nygaard, R., Nielbo, S., Schwartz, T. W. & Poulsen, F. M. (2006) The PP-fold solution structure of human polypeptide YY and human PYY3-36 as determined by NMR, *Biochemistry. 45*, 8350-7.
8. Mierke, D. F., Durr, H., Kessler, H. & Jung, G. (1992) Neuropeptide Y. Optimized solid-phase synthesis and conformational analysis in trifluoroethanol, *European Journal of Biochemistry. 206*, 39-48.
9. Bianchi, E., Rampone, R., Tealdi, A. & Ciferri, A. (1970) The role of aliphatic alcohols on the stability of collagen and tropocollagen, *J. Biol. Chem. 245*, 3341-5.
10. Herskovits, T. T., Gadegbeku, B. & Jaillet, H. (1970) On the structural stability and solvent denaturation of proteins. I. Denaturation by the alcohols and glycols, *J. Biol. Chem. 245*, 2588-98.
11. Kamatari, Y. O., Konno, T., Kataoka, M. & Akasaka, K. (1996) The methanol-induced globular and expanded denatured states of cytochrome c: a study by CD fluorescence, NMR and small-angle X-ray scattering, *J. Mol. Biol. 259*, 512-23.
12. de Jongh, H. H., Killian, J. A. & de Kruijff, B. (1992) A water-lipid interface induces a highly dynamic folded state in apocytochrome c and cytochrome c, which may represent a common folding intermediate, *Biochemistry. 31*, 1636-43.
13. Babu, K. R., Moradian, A. & Douglas, D. J. (2001) The methanol-induced conformational transitions of beta-lactoglobulin, cytochrome c, and ubiquitin at low pH: a study by electrospray ionization mass spectrometry, *J. Am. Soc. Mass. Spectrom. 12*, 317-28.
14. Wang, Y.-F., Ho, M.-Y. & Ho, Y.-P. (2004) Using mass spectrometry to probe subtle differences in conformations of several cytochromes c in aqueous and methanol solutions, *J. Mass Spectrom. 39*, 1523-1530.
15. Bychkova, V. E., Dujsekina, A. E., Klenin, S. I., Tiktopulo, E. I., Uversky, V. N. & Ptitsyn, O. B. (1996) Molten globule-like state of cytochrome c under conditions simulating those near the membrane surface, *Biochemistry. 35*, 6058-63.

16. Brutscher, B., Bruschweiler, R. & Ernst, R. R. (1997) Backbone dynamics and structural characterization of the partially folded A state of ubiquitin by 1H, 13C, and 15N nuclear magnetic resonance spectroscopy, *Biochemistry Biochemistry. 36*, 13043-53.

17. Harding, M. M., Williams, D. H. & Woolfson, D. N. (1991) Characterization of a partially denatured state of a protein by two-dimensional NMR: reduction of the hydrophobic interactions in ubiquitin, *Biochemistry. 30*, 3120-8.

18. Nakano, T. & Fink, A. L. (1990) The folding of staphylococcal nuclease in the presence of methanol or guanidine thiocyanate, *J. Biol. Chem. 265*, 12356-62.

19. Shiraki, K., Nishikawa, K. & Goto, Y. (1995) Trifluoroethanol-induced stabilization of the alpha-helical structure of beta-lactoglobulin: implication for non-hierarchical protein folding, *J. Mol. Biol. 245*, 180-94.

20. Alonso, D. O. & Daggett, V. (1995) Molecular dynamics simulations of protein unfolding and limited refolding: characterization of partially unfolded states of ubiquitin in 60% methanol and in water, *J. Mol. Biol. 247*, 501-20.

21. Kabsch, W. & Sander, C. (1983) Dictionary of Protein Secondary Strucutre: Pattern Recognition of Hydrogen-Bonded and Geometrical Features, *Biopolymers. 22*, 2577-2637.

22. Luginbuhl, P., Szyperski, T. & Wuthrich, K. (1995) Statistical Basis For the Use Of C-13-Alpha Chemical Shifts In Protein Structure Determination, *Journal of Magnetic Resonance Series B. 109*, 229-233.

23. Wishart, D., Sykes, B. & Richards, F. (1991) Relationship between nuclear magnetic resonance chemical shift and protein secondary structure., *J Mol Biol. 222*, 311-33.

24. Wishart, D. S. & Sykes, B. D. (1994) Chemical shifts as a tool for structure determination, *Methods Enzymol. 239*, 363-92.

25. Spera, S., Ikura, M. & Bax, A. (1991) Measurement of the exchange rates of rapidly exchanging amide protons: application to the study of calmodulin and its complex with a myosin light chain kinase fragment, *J Biomol NMR. 1*, 155-65.

26. Ferguson, N., Sharpe, T. D., Johnson, C. M., Schartau, P. J. & Fersht, A. R. (2007) Structural biology: analysis of 'downhill' protein folding, *Nature. 445*, E14-5; discussion E17-8.

27. Sadqi, M., Fushman, D. & Munoz, V. (2007) Structural BiologyAnalysis of 'downhill' protein folding; Analysis of protein-folding cooperativity (Reply), *Nature. 445*, E17-18.

28. Vamvaca, K., Vogeli, B., Kast, P., Pervushin, K. & Hilvert, D. (2004) An enzymatic molten globule: Efficient coupling of folding and catalysis, *P Natl Acad Sci Usa. 101*, 12860-12864.

29. Bader, R., Bettio, A., Beck-Sickinger, A. G. & Zerbe, O. (2001) Structure and Dynamics of Micelle-bound Neuropeptide Y: Comparison with Unligated NPY and Implications for Receptor Selection, *J. Mol. Biol. 305*, 307-329.

30. Tjandra, N. & Bax, A. (1997) Direct measurement of distances and angles in biomolecules by NMR in a dilute liquid crystalline medium, *Science. 278*, 1111-4.

31. Tjandra, N. & Bax, A. (1997) Measurement of dipolar contributions to 1JCH splittings from magnetic-field dependence of J modulation in two-dimensional NMR spectra, *J Magn Reson. 124*, 512-5.

32. Goddard, T. D., Kneller, D. G. SPARKY 3, *University of California, San Francisco*.

33. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, Wiley-Interscience, New York.

34. Vuister, G. W. & Bax, A. (1992) Resolution Enhancement and Spectral Editing of Uniformly 13C-Enriched Proteins by Homonuclear Broadband 13C Decoupling, *J. Mag. Reson. 98*, 428.

35. Bax, A., Clore, M. & Gronenborn, A. M. (1990) $^{1H\text{-}1H}$ Correlation via Isotropic Mixing of 13C Magnetization, a New Three-Dimensional Approach for Assigning 1H and 13C Spectra of 13C Enriched Proteins, *J. Magn. Reson. 88*, 425-431.

36. Güntert, P. (2004) Automated NMR structure calculation with CYANA, *Methods Mol. Biol. 278*, 353-78.

37. Cornell, W. D., Cieplak, P., Bayly, C. I., Gould, I. R., Merz, K. M., Ferguson, D. M., Spellmeyer, D. C., Fox, T., Caldwell, J. W. & Kollman, P. A. (1995) A Second Generation Force Field For the Simulation of Proteins, Nucleic Acids, and Organic Molecules, *J. Am. Chem. Soc. 118*, 2309.

38. Koradi, R., Billeter, M. & Wüthrich, K. (1996) MOLMOL: a program for display and analysis of macromolecular structures, *J Mol Graph. 14*, 51-5, 29-32.

39. Noggle, J. H. & Schirmer, R. E. (1971) *The Nuclear Overhauser Effect - Chemical Applications*, Academic Press, New York.

40. Meiboom, S. & Gill, D. (1958) Modified spin-echo method for measuring spin-relaxation rates, *Rev. Sci. Instrum. 29*, 688-691.

41. Press, W. H., Teukolsky, S. A., Vetterling, W. T. & Flannery, B. P. (1992) *Numerical Recipes. The Art of Scientific Computing*, 2nd edn, Cambridge University Press, Cambridge.

6.4 Supplementary material

Table S1: Chemical Shifts of pPYY, 1mM, d^3-MeOH, 301K, referenced to residual CD_2**H**OH at 3.31 ppm

N	Res.	H^N	H'^a	H^b	others
1	Tyr	-	-	-, -	δH -, -; εH -, -; ηOH -
2	Pro		4.43	1.97, 2.22	γCH_2 -, -; δCH_2 -, -
3	Ala	8.29	4.32	1.38	
4	Lys	8.16	4.63	1.86, 1.86	γCH_2 1.69, 1.69; δCH_2 1.51, 1.51; εCH_2 -, -; ζNH_3^+ -
5	Pro		4.49	-, -	γCH_2 2.09, 2.09; δCH_2 3.65, 3.80
6	Glu	8.29	4.36	1.94, 1.94	γCH_2 2.11, 2.43; εH -
7	Ala	8.07	4.63	1.34	
8	Pro		4.35	1.97, 2.23	γCH_2 -, -; δCH_2 -, -
9	Gly	8.52	3.80, 3.92		
10	Glu	8.07	4.29	2.42, 2.42	γCH_2 2.07, 2.15; εH -
11	Asp	8.27	4.67	2.79, 2.89	δH -
12	Ala	7.95	4.49	1.39	
13	Ser	8.13	4.7	3.98, 4.26	γOH -
14	Pro		4.3	1.94, 2.05	γCH_2 1.69, 1.69; δCH_2 -, -
15	Glu	8.24	4.09	2.21, 2.21	γCH_2 2.51, 2.59; εH -
16	Glu	7.96	4.11	2.17, 2.28	γCH_2 2.51, 2.51; εH -
17	Leu	8.34	4.11	1.86, 1.86	γH 1.50; δCH_3 0.93, 0.93
18	Ser	8.34	4.25	4.01, 4.10	γOH -
19	Arg	7.97	4.06	1.99, 1.99	γCH_2 1.67, 1.67; δCH_2 3.21, 3.21; εNH 7.45; ηNH_2 -, -
20	Tyr	8.44	4.17	3.13, 3.18	δH 6.76, 6.76; εH 6.56, 6.56; ηOH -
21	Tyr	8.67	4.13	3.12, 3.12	δH 7.16, 7.16; εH 6.77, 6.77; ηOH -
22	Ala	8.58	4.05	1.57	
23	Ser	8.23	4.17	4.03, 4.06	γOH -
24	Leu	8.35	4.1	1.68, 1.68	γH 1.50; δCH_3 0.87, 0.87
25	Arg	8.49	3.96	1.99, 1.99	γCH_2 -, -; δCH_2 3.20, 3.20; εNH 7.44; ηNH_2 -, -
26	His	8.4	4.48	3.41, 3.51	δ^1NH -; δ^2H 7.30; ε^1H 8.73; ε^2NH -

27	Tyr	8.49	4.07	3.21, 3.21	δH 6.97, 6.97; εH 6.64, 6.64; ηOH -
28	Leu	8.7	3.92	1.97, 1.97	γH 1.50; δCH$_3$ 0.93, 0.93
29	Asn	8.55	4.38	2.75, 3.10	δNH$_2$ 7.02, 7.79
30	Leu	8.01	3.97	1.74, 1.74	γH 1.60; δCH$_3$ 0.84, 0.84
31	Val	8.27	3.65	1.98	γCH$_3$ 0.75, 0.84
32	Thr	8.03	3.87	4.36	γCH$_3$ 1.28; γOH -
33	Arg	8.03	4.06	1.98, 1.98	γCH$_2$ 1.86, 1.86; δCH$_2$ 3.16, 3.16; εNH 7.44; ηNH$_2$ -, -
34	Gln	8.15	4.02	2.27, 2.27	γCH$_2$ 2.07, 2.52; εNH$_2$ -, -
35	Arg	7.89	4.07	1.65, 1.75	γCH$_2$ 1.43, 1.52; δCH$_2$ 3.01, 3.07; εNH 7.25; ηNH$_2$ -, -
36	Tyr	7.77	4.52	2.84, 3.20	δH 7.22, 7.22; εH 6.67, 6.67; ηOH -

Figure S2: Sequence plot of restraints used during the structure calculation of pPYY in MeOH

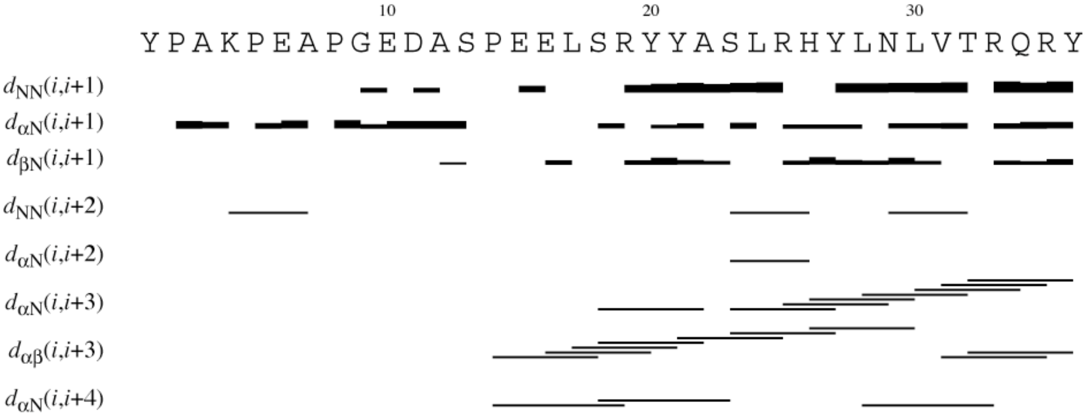


Table S3: Information on the structure calculation of pPYY in MeOH

Distance restraints	Total	229
	Intra-residual	90
	Sequential (i - j = 1)	76
	Medium (i - j = 2, 3, 4)	63
	Long-range	0
Dihedral angle restraints		129
RMSD[a] (Å)[b]	Tyr1-Tyr36 backbone	6.23 ± 1.32
	Tyr1-Tyr36 all heavy atoms	7.63 ± 1.44
	Tyr1-Glu15 backbone	4.82 ± 1.02
	Tyr1-Glu15 all heavy atoms	6.15 ± 1.14
	Pro14-Arg35 backbone	1.73 ± 1.14
	Pro14-Arg35 all heavy atoms	2.72 ± 1.00
	Glu16-Arg35 backbone	1.50 ± 1.00
	Glu16-Arg35 all heavy atoms	2.53 ± 0.88

[a] Atomic root mean square deviation calculated by superimposing backbone atoms of residues 17 to 31 (17 to 31) of the 20 minimized structures referenced to the mean coordinates.

[b] N, Cα, C´ atoms.

Table S7: $^3J_{HN\alpha}$ scalar coupling constants for pPYY in MeOH and in water. Values ≤ 6.0 Hz are shaded in grey.

Res.	pPYY in MeOH	pPYY
1		
2		
3	5.9 ± 0.1	6.1 ± 0.2
4	7.1 ± 0.1	5.3 ± 0.3
5		
6	7.7 ± 0.1	7.1 ± 0.2
7	7.6 ± 0.3	5.1 ± 0.3
8		
9	6.1 ± 0.8	7.4 ± 0.7
10	6.2 ± 0.2	5.5 ± 0.3
11	7.1 ± 0.1	9.8 ± 0.1
12	6.6 ± 0.1	5.0 ± 1.3
13	6.4 ± 0.3	5.7 ± 0.4
14		
15	5.7 ± 0.3	5.1 ± 0.3
16	6.1 ± 0.2	6.0 ± 0.5
17	5.3 ± 0.2	5.0 ± 0.4
18	5.0 ± 0.2	4.7 ± 0.5
19	5.2 ± 0.2	6.2 ± 0.2
20	5.0 ± 0.2	5.4 ± 0.8
21	4.8 ± 0.2	5.3 ± 1.0
22	4.9 ± 0.2	4.9 ± 0.3
23	5.1 ± 0.2	6.0 ± 1.4
24	5.0 ± 0.2	6.0 ± 0.2
25	5.0 ± 0.2	5.3 ± 0.9
26	4.9 ± 0.2	5.6 ± 0.2
27	4.9 ± 0.2	6.1 ± 0.4
28	4.7 ± 0.2	4.4 ± 1.2
29	5.0 ± 0.2	5.6 ± 0.4
30	5.1 ± 0.2	6.3 ± 0.3
31	5.0 ± 0.2	6.3 ± 0.1
32	5.1 ± 0.2	6.8 ± 0.3
33	5.2 ± 0.2	6.7 ± 0.3
34	5.3 ± 0.2	6.6 ± 0.2
35	6.0 ± 0.2	6.8 ± 0.2
36	9.3 ± 0.1	8.9 ± 0.2

7. Towards a model for cell-surface exposed carbohydrate moieties suitable for structural studies by NMR

In the present study a synthetic glycolipid system is presented that can be readily incorporated into phospholipid micelles, and which allows the study of cell-surface exposed carbohydrate units by high-resolution NMR techniques. Here, we present an efficient route for the synthesis of glycolipid compounds containing mannose, mannobiose or mannotriose coupled either directly to an alkyl chain or via a polyethylene glycol linker. Furthermore, we have validated our model system by measuring binding of cyanovirin-N (CV-N), a cyanobacterial protein that binds with nanomolar affinity to the terminal arms of high-mannose structures of the HIV surface envelope glycoprotein gp120, to glycolipids whose carbohydrate portion comprises the corresponding high-mannose moieties. From results of chemical shift mapping utilizing uniformly ^{15}N-labelled CV-N we conclude that binding to the protein occurs at sites similar to those involved in binding the non-conjugated carbohydrates. We characterized the insertion of the glycolipids into the DPC micelles by measurements of translational diffusion and observed that the diffusion constants of the glycolipids are very similar to those of the DPC micelles themselves but significantly deviate from those of the free glycolipids. We also present experimental proof that the glycolipids remain inserted in the micelles while binding to CV-N. Finally, by adding a ligand with higher affinity to CV-N, but which is not coupled to a lipid anchor, CV-N may be released from the glycolipid and hence from the micelle-associated state.

published as: Mareš, J., Müller, J. U., Skirgailiene, A., Neumoin, A., Bewley, C.A., Schmidt, R. R., and Zerbe, O. (2006) *ChemBioChem.* 7, 1764-1773.

Introduction

Cell-cell recognition is an event of prime biological importance in a variety of biological phenomena such as cell migration, organ formation, immune defense and microbial infection[1-3]. Specificity of the interaction is mediated by various tissue-dependent, receptors. Much attention has been recently focused on recognition processes which are mediated by interactions of membrane-coupled carbohydrates with lectins[1-4].Although extensive biochemical work has been done in that field to determine which units of the oligosaccharides are important for the recognition process, little structural data on their interaction with proteins are available and not much is known about the influence of the membrane environment.

NMR has now been established as an important method for deriving high-resolution structural data of proteins in solution[5-8]. Whereas the overwhelming majority of molecules under study are cytosolic proteins smaller than 20 kDa the potential of NMR for studies of membrane proteins or membrane-associated peptides is increasingly recognized[9-13] despite technological challenges due to technical problems associated with this class of molecule[14, 15]. Membrane models frequently used are phospholipid detergents[11, 16-18] such as the phosphatidyl derivatives or dodecylphosphocholine (DPC)[19], or sodiumdodecylsulfate (SDS). Whereas at least the former seem to reasonably represent the chemical environment of the phospholipids commonly encountered on the surfaces of biological membranes such models largely neglect the chemical diversity of other components that populate biological membranes. In this work we present a model for cell-surface exposed carbohydrate units that readily integrates into common membrane models, in particular into phospholipid micelles, and which allows the study of carbohydrate-protein interactions in detail by high-resolution NMR methods.

A model for such a mimic of cell-surface exposed carbohydrate units should ideally have the following properties: i) it should readily integrate into established membrane models, in particular into the phospholipid micelles commonly used in high-resolution NMR studies, ii) it must be readily accessible synthetically, simple in its architecture while permitting expansion of its complexity, and iii) it should be recognized by naturally occurring carbohydrate-binding proteins. Herein, we present the design of such a system using the 1-3 or D1 arm[20] of oligomannose as our model. As seen in Figure A, the model consists of a carbohydrate portion, a lipid chain, and if desired a polyethylene glycol spacer connecting the two. The lipid chain allows insertion into the phospholipid micelles and can be tuned to be compatible with the desired phospholipid system by varying the lipid chain length (C12 in our case since we use dodecylphosphocholine (DPC) as the detergent). The flexible ethylene glycol spacer, whose length can also be varied, allows the carbohydrate unit to adjust its position relative to the micelle surface, and to traverse the gradient in lipophilicity across the head group region of the micelles. Finally, the spacer is connected to the carbohydrate portion of interest.

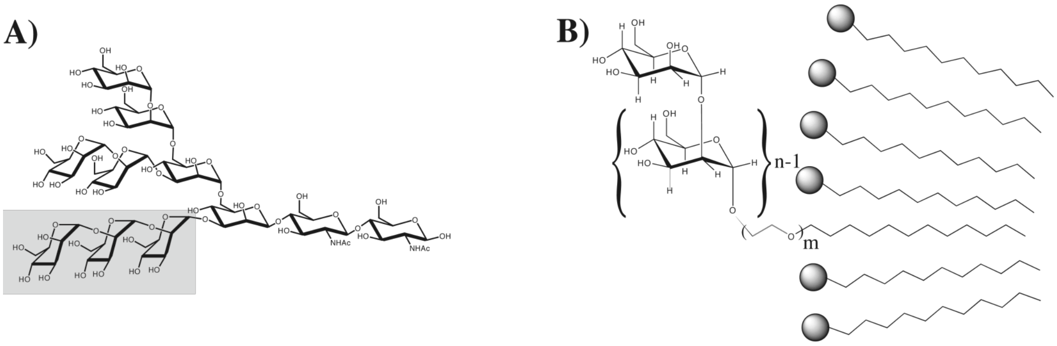


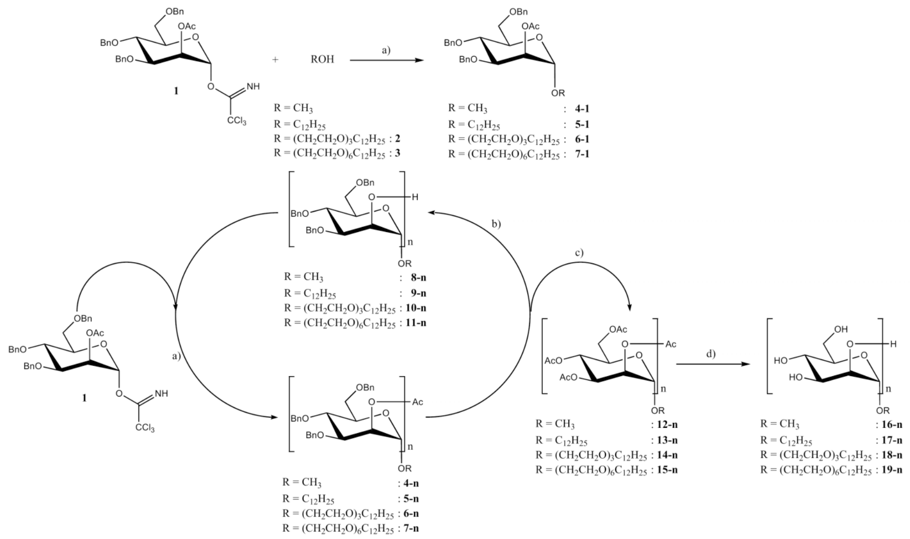
Figure A: Model of membrane-bound carbohydrates. (**a**) Chemical structure of high-mannose-type N-linked glycans. The segment that is mimicked by the model has been shaded in grey. (**b**) Chemical structure of the model compounds representing the D1 arm of the high-mannose structure. Balls and chains denote the phospholipid headgroups and lipid chains of the DPC molecules, respectively.

In this article we describe the synthesis of the glycolipids and present NMR data to prove that the glycolipids are properly inserted into the micelles. We then describe the recognition of mannose-derived glycolipids by cyanovirin N (CV-N), a cyanobacterial protein that binds to the high-mannose part of the HIV viral envelope protein gp120[21], as the proof-of-principle. We will demonstrate that derivatives of 1,2-α-mannobiose and 1,2-α-mannotriose are properly recognized by CV-N. Moreover, the glycolipids remain inserted in the micelles when bound to the protein. Finally, we present a way of replacing glycolipid bound to CV-N thereby releasing the protein from the micelle-bound state again.

7.1 Results

Synthesis of the glycolipids

The target oligomannosides were synthesized with the trichloroacetimidate **1**[22, 23] as donor; **1** has already been used in studies on the solution synthesis[22] as well as the solid phase[24] synthesis of oligomannosides. Due to neighbouring group participation, the 2-*O*-acetyl group will lead to α-product formation; additionally, the *O*-acetyl group provides the required temporary protection for the desired chain extension. Starting with the aglycon moieties methanol, dodecanol, **2** or **3** glycosylation and deprotection steps were employed in a cyclic manner (Scheme 2). For reference purposes methyl mannosides were prepared as well.



Scheme 2. Synthetic route to the glycolipids. The following compounds were synthesized: **1**, **2**, **3**, **4-1**, **4-2**, **4-3**, **5-1**, **5-2**, **6-1**, **6-2**, **7-1**, **7-2**, **7-3**, **7-4**, **7-5**, **12-2**, **12-3**, **13-2**, **14-2**, **15-1**, **15-2**, **15-3**, **15-4**, **15-5**, **16-2**, **16-3**, **17-2**, **18-2**, **19-1**, **19-2**, **19-3**, **19-4**, **19-5**. Reagents used are **a**) TMSOTf, CH_2Cl_2; **b**) NaOMe, MeOH; **c**) **i**. H_2, $Pd(OH)_2/C$, THF/MeOH; **ii**. Ac_2O, pyr; **d**) NaOMe, MeOH.

For the glycosylation steps, trimethylsilyl trifluoromethanesulfonate (TMSOTf) was used as catalyst (0.1 equiv with regard to the donor **1**), yielding the fully protected mannosides **4-n**, **5-n**, **6-n** and **7-n**. The formation of orthoester by-products was not observed under these conditions.

The deacetylation steps were carried out under standard conditions (NaOMe in MeOH) yielding the 2-*O*-deprotected compounds **8-n**, **9-n**, **10-n** and **11-n**. These were used in the following glycosylation step without further purification.

To obtain the final substances portions of the benzyl-protected compounds (**4-n**, **5-n**, **6-n** and **7-n**) were removed from the elongation cycle. To simplify purification of the glycolipids or oligomannosides the benzyl protecting groups were removed and the products were per-*O*-acetylated using a two-step procedure to furnish **12-n**, **13-n**, **14-n** and **15-n**. Upon *O*-deacetylation the desired mannosides **16-n**, **17-n**, **18-n** and **19-n** are obtained in high purity with no need for further purification.

Incorporation of the glycolipids into phospholipid micelles

Glycolipids themselves are amphiphilic molecules and therefore are expected to assemble into micelles on their own. Accordingly, we haven chosen the phospholipid-to-glycolipid ratio to be very large (usually well-above 300:1) to facilitate insertion of the glycolipids into the phospholipid micelles without forming glycolipid clusters in the micelles[13]. To experimentally demonstrate that glycolipid clusters are not formed we have compared velocities for translational diffusion of the glycolipids measured in the absence and in the presence of phospholipid micelles. To this end we use a stimulated pulsed field gradient echo sequence, similarly to a methodology applied to study peptide-micelle interactions[25]. The diffusion coefficients of the derivatives of mannobiose and mannotriose as well as for DPC are presented in Table 1. The data clearly demonstrate that the glycolipids diffuse at rates similar to those for unloaded DPC micelles and significantly slower than in the absence of DPC micelles indicating that they are incorporated into the phospholipid micelles:

	17-2	**18-2**	**19-2**	**19-3**	DPC
Water	$1.7{\pm}0.05 \bullet 10^{-10}$	$1.3{\pm}0.02 \bullet 10^{-10}$	$1.1{\pm}0.03 \bullet 10^{-10}$	$1.5{\pm}0.02 \bullet 10^{-10}$	$7.6{\pm}0.1 \bullet 10^{-10}$
DPC	$7.4{\pm}0.5 \bullet 10^{-11}$	$7.6{\pm}0.5 \bullet 10^{-11}$	$7.7{\pm}0.4 \bullet 10^{-11}$	$7.2{\pm}0.06 \bullet 10^{-11}$	

Table 1: Diffusion constants in m^2/s for glycolipids in water or DPC micelle solution as measured by pulsed field gradient NMR

To further characterize the glycolipid/phospholipids micelles, we introduced the micelle-integrating spin label 5-doxylstearate into the phospholipid solution containing **17-2**. The paramagnetic doxyl moiety of the spin label is positioned close to the phospholipid head groups[9] and through very efficient electron-nuclear spin dipolar interaction enhances transverse relaxation of all protons within a 10-15 Å radius. We could measure significant reductions in signal intensities for protons of those mannose residues that are directly coupled to the lipid chain, and much less decay for signals derived from protons of the terminal mannose unit. Given the precision of the spin-label data it is impossible to derive accurate positions of the mannose units relative to the water-membrane interface. Nevertheless, the data clearly indicate that protons from the first mannose unit are closer to the interface on average, and that the protons corresponding to positions at C-3' to C-6' of the terminal mannose are pointing away from the interface (see Supp. Mat.). We have conducted similar experiments using **19-2**, that additionally contains the hexaethyleneglycol spacer. To summarize, spinlabel effects were smaller and the differences between attenuations of the protons of the two mannose rings is less pronounced. This is in agreement with the expectation that due to the extended glycol linker additional flexibility is introduced. To conclude, the spin label data prove that the glycolipids are integrated into the micelles and thereby supports the conclusions from the diffusion data. However, it is likely that residual flexibility about the glycosidic bonds remains.

Recognition of the model by cyanovirin N

To validate our system we have measured binding of cyanovirin N[21], a cyanobacterial carbohydrate binding protein that specifically recognizes α-linked oligomannose, to carbohydrate moieties containing 1,2-α-linked mannose units. Cyanovirin N (CV-N) is undergoing preclinical studies as a potential drug for anti-HIV treatment. Entry of the HIV viral genome into mammalian cells is triggered through binding of the HIV viral envelope protein gp120 to the mammalian cell surface receptor CD4, resulting in insertion of a fusion peptide from gp41 into the mammalian membrane and which, via a spring-loaded mechanism, acts to pull the two membranes closer together resulting in fusion of the viral and the mammalian membranes[26-28]. It was recognized that CV-N binds to the high-mannose part of glycosylated residues from gp120 thereby inhibiting its interaction with CD4 and preventing the above-described cascade of events that lead to fusion of the membranes[26-28]. CV-N has two oligosaccharide binding sites. Its structure has been determined both crystallographically[29] and by NMR methods[30, 31] in the absence and in the presence of carbohydrate ligands and proved to exhibit some features in common with those of other lectins. It is also known that CV-N binds to 1,2-α-linked mannobiose via two binding sites, although with different affinities with K_a values of $7.2\pm4 \times 10^6$ M^{-1} and $6.8\pm4 \times 10^5$ M^{-1} for the high- and low-affinity binding sites, respectively[32]. In addition, thermodynamic parameters of CV-N binding to trimannosides corresponding to the three different arms of the high-mannose part have been determined by isothermal titration calorimetry (ITC)[33]. That study demonstrated that the highest affinity was observed for 1,2-α-linked mannotriose with K_a values of $6.6\pm0.7\times10^6$ M^{-1} and $3.7\pm0.3\times10^5$ M^{-1} for the high- and low-affinity binding sites, respectively. Binding constants for association of CV-N to trimannosides corresponding to the other two arms of the high-mannose portion were generally at least one order of magnitude lower. Moreover, Sandstrom *et al.* have used saturation transfer difference techniques to establish the mode by which carbohydrates bind to CV-N[34].

Building onto these results we decided to synthesize glycolipids corresponding to 1,2-linked α-mannose. We have kept the design of the lipid chain unmodified corresponding to a C-12 chain, but varied the number of ethylene glycol units in the flexible spacer. Initial screens for binding to CV-N were performed with derivatives of the 1,2-linked α-mannobiose. Constructs tested comprised glycolipids with none (**17-2**), three (**18-2**) and six (**19-1**, **19-2**, **19-3**, **19-4** and **19-5**) ethyleneglycol units as the spacer.

We have tested binding of 1,2-linked α-mannobiose to CV-N via NMR chemical shift mapping methodology by measuring chemical shift changes of CV-N in the presence of glycolipid-loaded micelles relative to a reference spectrum recorded in the presence of DPC micelles alone. The results are shown in Figure 1, where residues whose backbone amide resonances undergo the largest changes in chemical shift upon binding to mannobiose (**19-2**) and mannotriose (**19-3**) glycolipds are depicted as grey spheres.

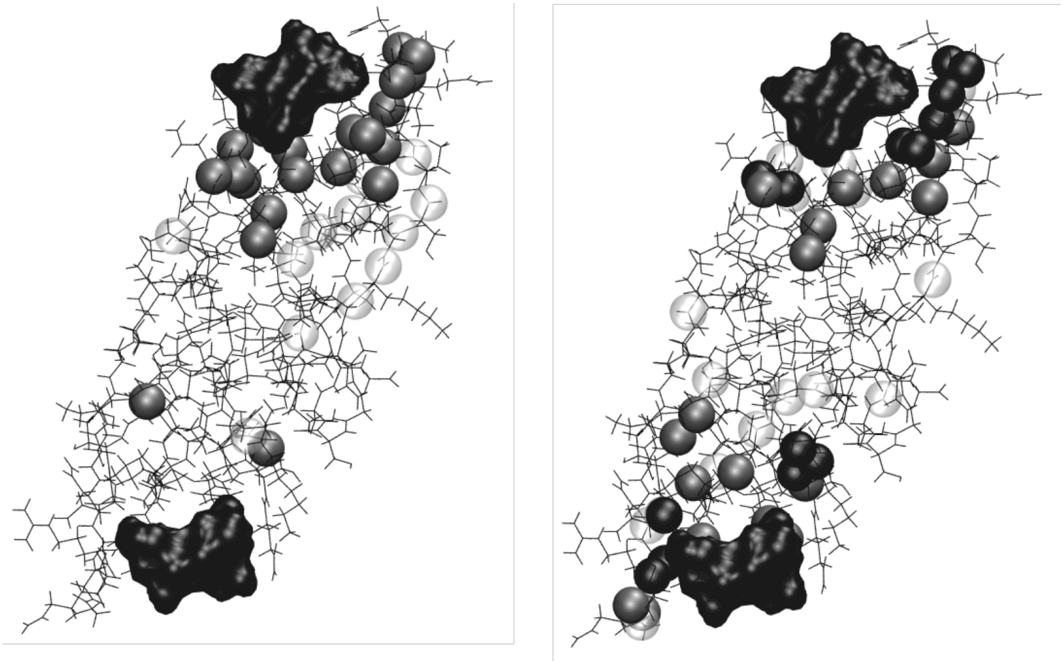


Figure 1: Results from chemical shift mapping experiments for binding of **19-2** (left) and **19-3** (right) to CV-N. The carbohydrate binding surfaces of α-1,2-mannobiose as extracted from the structure of Bewley[30] is depicted as a black solid surface.

Most of the changes occur for residues close to the high-affinity binding site. However, shift changes are also observed at slightly more remote regions suggesting that smaller structural adjustments occur upon binding. In case of the glycolipid derived from the trimannoside **19-3** shift changes occur in the vicinity of both carbohydrate-binding sites. Again, additional changes at more remote positions are observed. In the case of both mannobiose **19-2** and mannotriose **19-3** doubling of peaks corresponding to resonances of free and bound CV-N occurred when substochiometric amounts of glycolipid were added indicating that the complex is in slow exchange on the NMR timescale. We have also synthesized the corresponding tetra- and pentamannoside glycolipid derivatives **19-4** and **19-5**. We observed very similar changes in the chemical shift mapping experiments, again restricted to the vicinity of the two binding sites. In both cases larger changes in peak positions in the ^{15}N-^{1}H correlation map when compared to **19-3** are observed for residues 24-29 comprising the peptide segment ^{21}Thr-Cys-Glu-^{24}Arg, which constitute the rim of one of the two binding pockets. We suspect that the additional mannose units are making contacts with those residues, and work to characterize that interaction in more detail is presently in progress. A similar situation occurs for residues 78-79, which are part of a loop forming the rim of the other binding pocket.

We have also noticed a dramatic decrease of signal intensity of CV-N resonances upon adding the mannotriose **19-3**. In principle, two mechanisms may account for this observation: i) binding of the mannotriose takes place via both the low- and the high-affinity binding site. Since the large excess of phospholipid ensures that micelles are only occupied by a single glycolipid the high K_d of mannotriose for both sites suggests that CV-N binds to two different, glycolipid-loaded micelles simultaneously. The resulting dramatic increase in molecular weight to above 50kD leads to very efficient transverse relaxation accompanied by substantial signal loss in the 2D spectra. ii) Tight coupling of CV-N to membranes may lead to partial denaturing of the protein, which may possibly be accompanied by broadening due to exchanging conformers. To test whether CV-N may be unfolded when forced into the membrane interface we have dissolved CV-N in a 50:50 water:methanol mixture, a solvent that in our experience reproduces properties of the micelle-water interface quite well. In such a solvent CV-N clearly converts into a molten-globule state (data not shown). Thus, to reveal which of the two mechanisms leads to the observed signal loss we expressed a triple amino acid mutant of CV-N in which the low-affinity binding site was knocked out[35]. The data of the shift-mapping experiments for the mutant are compared to those derived from the wild-type protein in Figure 2.

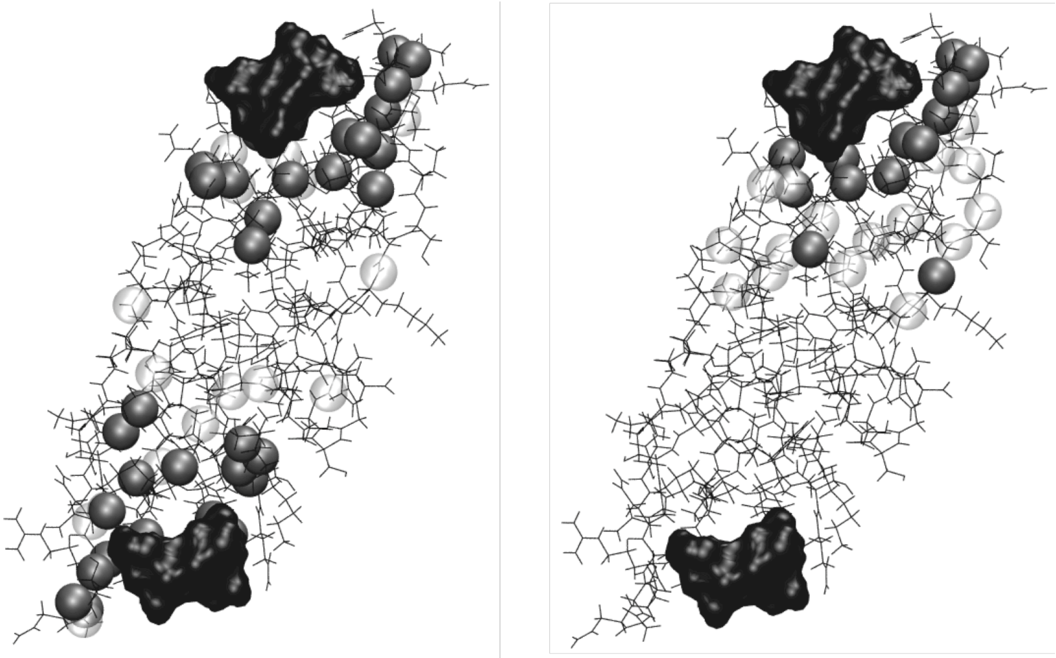


Figure 2: Results from chemical shift mapping experiments for binding of **19-3** to CV-N (left) and the K3N/E23I/N93A mutant CV-N, in which one binding site has been knocked out (right).

Clearly, significant shift changes are limited to the direct vicinity of the remaining binding site, which reveals that allosteric effects are probably less important in this case. Moreover, signal loss in the HSQC spectra is small and comparable to that occurring upon addition of the mannobiose **19-2** indicating that coupling of two mannotriose-loaded micelles to intact CV-N accounts for the observed loss in signal.

The average distance by which the carbohydrate units are separated from the micelle surface may have an influence on how well they can bind to their target. Accordingly, we have varied the length of the spacer unit and tested binding of CV-N to **17-2** (no spacer), **18-2** (triethyleneglycol spacer) and **19-2** (hexaethyleneglycol spacer). All of these glycolipids bind to CV-N, and chemical shift changes occurred at identical positions. In addition, exchange between the free and all of the glycolipid-bound species is slow on the NMR timescale indicating that binding affinities are not dramatically affected. We have noticed that the induced chemical shift changes are not identical in case of the three glycolipids. When comparing the shift changes induced upon adding **17-2** with those induced by **19-2** we observed highly similar changes for all positions that are not surface exposed. In contrast, larger changes occur for amide moieties of the loop encompassing residues 74 to 79 formed by the segment ^{74}Lys-Thr-Arg-Ala-Gln-^{79}Gln (see Supp. Mat.). This may possibly caused by the vicinity of these residues to the phospholipid headgroups in the micelle-bound state or the additional presence of oxygen atoms in the linker of **19-2**. Control experiments conducted in the absence of DPC indicate that the chemical shift changes in the loop comprising residues 77 to 79 upon adding **17-2** are largely due to the presence of DPC (data not shown). We have again added the spin label 5-doxylstearate to probe for vicinity of CV-N to the head group region. Attenuations in the case of CV-N binding to **17-2** are observed for residues close to the binding site and exposed to the surface of CV-N, and are most pronounced for residues of the loop encompassing residues 74-79 (see Supp. Mat). In addition, smaller changes occur for the aromatic, surface-exposed residues, for which favorable partitioning into the head group region could be demonstrated previously[36, 37]. In case of **19-2** spin label effects on signal resonances of CV-N were very much weaker and essentially limited to Tyr-29.

An important point concerns the question whether the glycolipids remain integrated in the micelles while interacting with CV-N or whether they are pulled out of the micelle. We have demonstrated binding of the glycolipids to CV-N and insertion of the glycolipids in the micelles, and the spin label data indicate that the glycolipid-bound CV-N is micelle-anchored. To gain further evidence for the latter we have chosen the following strategy to investigate this point, which is depicted in Figure 3.

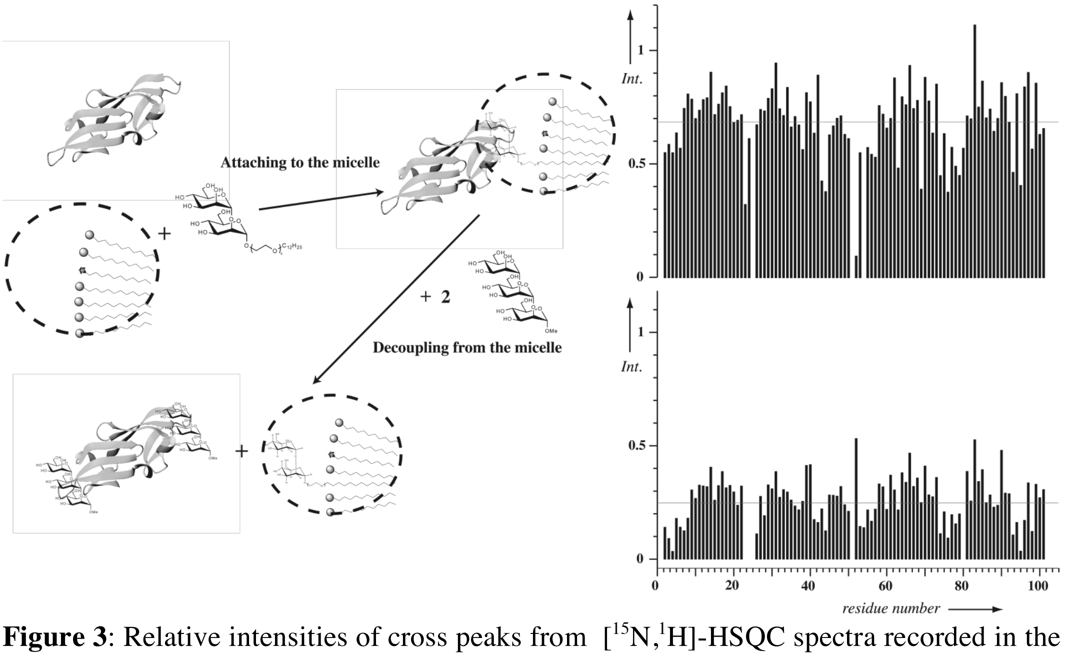


Figure 3: Relative intensities of cross peaks from [^{15}N,^{1}H]-HSQC spectra recorded in the presence of **19-2** integrated into DPC micelles to those in the absence of the glycolipid (bottom right) and after adding the non-lipidated mannotriose **16-3** (top right). The average intensity ratio of all residues is indicated by a horizontal line, and is 0.27 in case of bound **19-2** *vs*. free CV-N and 0.69 in case of bound **16-3** *vs*. free CV-N.

Therein, we compare the relative intensity of cross peaks in the [^{15}N,^{1}H]-HSQC spectrum of CV-N in the presence of DPC micelles to those after adding **17-2**. Clearly, the signal loss upon addition of **19-2** is due to the enhanced transverse relaxation of the ternary complex formed by CV-N, the glycolipid and the DPC micelle. Upon addition of the *non-lipidated* mannotriose **16-3** the glycolipid anchor is replaced thereby decoupling the protein from the micelle, resulting in regain of signal intensity. To conclude, these data prove that CV-N is bound to micelle-imbedded **19-2**, and also indicate a way of decoupling the system from the micelle again. These data are additionally supported by measurements of ^{15}N relaxation rates. The overall correlation time τ_m of CVN in DPC solution, in the presence of the free dimannoside **16-2** as well as in the presence of **19-2** as computed from the R2/R1 ratio of ^{15}N nuclei from residues within secondary structural elements is 5.7ns, 5.8ns and 11.6ns, respectively (data not shown). This large change in the overall correlation time in the presence of **19-2** indicates that the observed line-broadening is due to formation of a complex of almost twice the molecular weight, which is only possible if the glycolipid is still anchored within the phospholipids micelle.

7.2 Conclusions

Herein we have presented a model for structural studies of cell-membrane surface exposed carbohydrate units. We have shown that the necessary glycolipids can be synthesized in an efficient manner, that they readily integrate into membrane models commonly used in structural NMR studies, and that they are properly recognized. Such models may enable the study of protein-carbohydrate interactions using systems that are synthetically more easily accessible than the complex glycopeptides or glycolipids found in nature. Incorporation of these glycolipids also adds complexity to the commonly used membrane models and thereby allows studying membrane-associated peptides or proteins or even integral membrane proteins in an environment that better mimics a biological membrane. Last but not least we have been able to demonstrate that binding to the glycolipid and hence coupling to the membrane can be reversed upon adding a high-affinity ligand. This may enable a route in which (preferably smaller) lectin domains are fused to the protein of interest in order to force a protein with little affinity to the membrane into its vicinity in a reversible manner.

7.3 Experimental Section

General methods. Solvents were purified by distillation and dried by standard procedures. The boiling range of the petroleum ether used was 35-70 °C. Thin layer chromatography (TLC) was performed on E. MERCK Si gel 60 F_{254} plates (0.2 mm) and E. MERCK RP-18 F_{254} plates (0.2 mm). The plates were visualized by immersion in mostain (200 mL 10% H_2SO_4, 10 g $(NH_4)_6Mo_7O_{24} \cdot 4\ H_2O$, 200 mg $Ce(SO_4)_2$), 10% H_2SO_4 or $KMnO_4$ solution (1% in water, 1% $NaHCO_3$) followed by heating. Preparative flash chromatography was carried out on MACHEREY-NAGEL silica gel 60 (43 – 60 μm) at a pressure of 0.02 – 0.04 MPa. Preparative MPLC was performed with a continuous flow rate (10 mL min^{-1}, pressure 5–10 bar) on MERCK Si gel LiChroprep Si 60 (15-25 μm) with a column sized 28×2.5 cm. Detection was carried out with a differential refractometer from KNAUER. ^1H-NMR and ^{13}C-NMR spectra for glycolipid analytics were recorded on JEOL LA 400 or BRUKER DRX 600 instruments. Proton chemical shifts are reported in ppm relative to the signal from residual solvent protons or Me_4Si as an internal standard. Assignments of ^1H and ^{13}C chemical shifts were carried out with the aid of COSY, HMQC and ROESY experiments. The single sugar moieties of oligosaccharides are referred to by small letters beginning at the sugar´s reducing end. Measurements of optical rotations were performed on a PERKIN-ELMER polarimeter 241 MC (1 dm cell). MALDI-TOF MS were recorded on a KRATOS ANALYTICAL Kompact Maldi II instrument or a Bruker Biflex III spectrometer in positive, linear mode with a delayed extraction MALDI source and a pulsed nitrogen laser (337 nm) with 2,5-dihydroxybenzoic acid (DHB) or α-cyano-4-hydroxy-cinnamic acid (CHCA) as the matrix (positive mode). ESI-IT MS were performed on a Bruker Esquire 3000+ (positive mode). Elemental analyses were performed by the microanalytical facilities at the University of Konstanz.

Synthesis of glycolipids: In the following we will outline a general procedure for synthesizing the glycolipids (see also Scheme I). The synthesis comprises the following steps: i) glycosylation yielding fully protected mannosides, ii) selective deprotection of the hydroxyl function at C-2 of the terminal mannose moiety, followed by iii) introduction of the orthogonal acetyl protection group at OH (C-2), *O*-debenzylation and subsequent complete *O*-acetylation, iv) and finally complete O-deacetylation.

General procedure I (GP I): Glycosylation with glycosyl donor 1 yielding fully protected mannosides

Under nitrogen atmosphere the glycosyl donor **1** (1.2-1.3 eq.) and the acceptor were dissolved in dry dichloromethane and TMSOTf (0.12-0.13 eq) was added. The reaction mixture was stirred at room temperature (30 min), neutralized with triethylamine and concentrated *in vacuo*. After neutralization with triethylamine the reaction mixture was concentrated under reduced pressure. The remaining solid was purified by column chromatography (see individual experimental details) to afford the fully protected mannosides. Depending on whether the acceptor is an alcohol or the suitably protected mannoside this step will yield the lipidated protected monosaccharide or the respective oligosaccharide entities.

General procedure II (GP II): Deacetylation

The protected compound was dissolved in MeOH and NaOMe was added. The mixture was stirred at RT until starting material was undetectable (TLC). Subsequent neutralization of the reaction mixture with ion-exchange resin (Amberlite IR-120 H^+ form) was followed by filtration. The filtrate was concentrated to obtain the deprotected compounds with yields greater than 90 %. Partially benzylated compounds were sufficiently pure for the following reactions and were used without further purification. Fully deprotected compounds were dissolved in water and subsequently lyophilized to afford the desired compounds for the NMR studies.

General procedure III (GP III): Replacement of benzyl protecting groups by acetyl groups

(i) A mixture of the benzylated precursor and Pearlman's catalyst in methanol/THF (1:1) was degassed and saturated with hydrogen several times, followed by vigorous stirring under hydrogen (12-16 hours). The catalyst was removed by filtration and concentrated *in vacuo* (RT) to yield partially deprotected compounds that were used without purification in the next step.

(ii) The products from step (i) were dissolved in pyridine/acetic anhydride (2:1) and stirred (12-16 hours). The reaction mixture was concentrated and the residue was purified by column chromatography (see individual experimental details). After purification the fully acetylated mannosides were obtained as foams.

The glycol linkers **2** and **3** were prepared from triethylene glycol and hexaethylene glycol, respectively, via monotritylation, alkylation with bromododecane and detritylation by standard procedures (see Supp. Mat. and references therein).

Synthesis of Dodecyl α-D-mannopyranosyl-(1→2)-α-D-mannopyranoside (17-2)

Compound **17-2** (43 mg, 84 μmol, 98%) was prepared from **13-2** (69 mg, 86 μmol) according to GP II as a colorless lyophilizate.

TLC: $R_f = 0.62$ (RP-18-silicagel, MeCN/H$_2$0 2:1); $[\alpha]_D^{20} = +48.1$ (c = 1.0, MeOH); ^{1}H-NMR (600 MHz, d_4-MeOD): δ = 0.88-0.92 (m, 3H, CH_3), 1.24-1.41 (m, 18H, 9CH_2), 1.56-1.61 (m, 2H, 1CH_2), 3.40-3.45 (m, 1H, 1/2O$CH_2C_{11}H_{23}$), 3.48-3.53 (m, 1H, *H*-5a or *H*-5b), 3.55-3.61 (m, 2H, *H*-4a, *H*-4b), 3.63-3.75 (m, 5H, *H*-3b, *H*-5a or *H*-5b, *H*-6a, *H*-6b, 1/2O$CH_2C_{11}H_{23}$), 3.79-3.86 (m, 4H, *H*-2a, *H*-3a, *H*-6´a, *H*-6´b), 3.98 (dd, 1H, *H*-2b), 4.96 (d, 1H, *H*-1b), 5.05 (d, 1H, *H*-1a) ppm; ^{13}C-NMR (151 MHz, d_4-MeOD): δ = 14.49 (1C, CH_3),23.77-33.11 (10C, 10CH_2), 63.08 (1C, *C*-6a or *C*-6b), 63.14 (1C, *C*-6a or *C*-6b), 68.72 (1C, O$CH_2C_{11}H_{23}$), 68.85 (1C, *C*-4a or *C*-4b), 69.07 (1C, *C*-4a or *C*-4b), 71.92 (1C, *C*-2b), 72.25 (1C, *C*-3a), 72.45 (1C, *C*-3b), 74.63 (1C, *C*-5a or *C*-5b), 75.01 (1C, *C*-5a or *C*-5b), 80.78 (1C, *C*-2a), 99.94 (1C, *C*-1a), 104.23 (1C, *C*-1b) ppm; elemental analysis is calcd. (%) for $C_{24}H_{46}O_{11}$ (510.6): C: 56.45, H: 9.08; found: C: 56.44, H: 9.30; MALDI-MS (pos. mode, CHCA): $[M+Na]^+$ calcd.: 533.3, found: 533.6; $[M+K]^+$ calcd.: 549.3, found: 549.5.

Synthesis of 3,6,9-Trioxahenicosanyl α-D-mannopyranosyl-(1→2)-α-D-mannopyranoside (18-2)

Compound **18-2** (35 mg, 54 μmol, 93%) was prepared from **14-2** (54 mg, 58 μmol) according to GP II as a colorless lyophilizate.

TLC: : R_f = 0.33 (RP-18-silicagel, MeCN/H_20 3:2); $[\alpha]_D^{20}$ = +38.5 (c = 1.0, MeOH); ^{1}H-NMR (600 MHz, d_4-MeOD): δ = 0.88-0.92 (m, 3H, C*H*$_3$), 1.24-1.41 (m, 18H, 9C*H*$_2$), 1.54-1.60 (m, 2H, 1C*H*$_2$), 3.45-3.49 (m, 2H, OC*H*$_2$$C_{11}H_{23}$), 3.53-3.75 (m, 18H, *H*-3b, *H*-4a, *H*-4b, *H*-5a, *H*-5b, *H*-6a, *H*-6b, 11/2OC*H*$_2$), 3.81-3.87 (m, 5H, *H*-2a, *H*-3a, *H*-6´a, *H*-6´b, 1/2OC*H*$_2$), 3.99 (dd, 1H, *H*-2b), 4.97 (d, 1H, *H*-1b), 5.11 (d, 1H, *H*-1a) ppm; ^{13}C-NMR (151 MHz, d_4-MeOD): δ = 14.50 (1C, *C*H$_3$), 23.74-33.08 (10C, 10*C*H$_2$), 63.03 (1C, *C*-6a or *C*-6b), 63.12 (1C, *C*-6a or *C*-6b), 67.89 (1C, 1O*C*H$_2$), 68.79 (1C, *C*-4a or *C*-4b), 68.99 (1C, *C*-4a or *C*-4b), 71.14-71.55 (5C, 5O*C*H$_2$), 71.86 (1C, *C*-2b), 72.08 (1C, *C*-3a), 72.39 (1C, *C*-3b or O*C*H$_2$$C_{11}H_{23}$), 72.41 (1C, *C*-3b or O*C*H$_2$$C_{11}H_{23}$), 74.62 (1C, *C*-5a or *C*-5b), 74.98 (1C, *C*-5a or *C*-5b), 80.57 (1C, *C*-2a), 100.12 (1C, *C*-1a), 104.18 (1C, *C*-1b) ppm; $C_{30}H_{58}O_{14}$ (642.8); MALDI-MS (pos. mode, CHCA): $[M+Na]^+$ calcd.: 665.4, found: 665.7; $[M+K]^+$ calcd.: 681.4, found: 681.7.

Synthesis of 3,6,9,12,15,18-Hexaoxatriacontanyl α-D-mannopyranoside (19-1)

Compound **19-1** (114 mg, 186 μmol) was prepared from **15-1** (152 mg, 195 μmol) according to GP II as a pale, colorless oil.

TLC: R_f = 0.24 (RP-18-silicagel, MeCN/H_20 2:1); $[\alpha]_D^{20}$ = +21.2 (c = 1.0, CHCl$_3$); ^{1}H-NMR (600 MHz, d_4-MeOD): δ = 0.89-0.92 (m, 3H, C*H*$_3$), 1.23-1.39 (m, 18H, 9C*H*$_2$), 1.54-1.60 (m, 2H, 1C*H*$_2$), 3.45-3.48 (m, 2H, OC*H*$_2$$C_{11}H_{23}$), 3.55-3.73 (m, 27H, *H*-3, *H*-4, *H*-5, *H*-6, 23/2OC*H*$_2$), 3.80-3.86 (m, 3H, *H*-2, *H*-6´, 1/2OC*H*$_2$), 4.80 (d, $^3J_{1,2}$ = 1.7 Hz, 1H, *H*-1) ppm; ^{13}C-NMR (151 MHz, d_4-MeOD): δ = 14.49 (1C, *C*H$_3$), 23.77-33.12 (10C, 10*C*H$_2$), 63.02 (1C, *C*-6), 67.84 (1C, 1O*C*H$_2$), 68.69 (1C, *C*-5), 71.21-71.68 (11C, 11O*C*H$_2$), 72.11 (1C, *C*-2), 72.43 (1C, O*C*H$_2$$C_{11}H_{23}$), 72.63 (1C, *C*-3), 74.66 (1C, *C*-4), 100.82 (1C, *C*-1) ppm; elemental analysis is calcd. (%) for $C_{30}H_{60}O_{12}$ (612.8): C: 58.80, H: 9.87; found: C: 58.25, H: 9.34; MALDI-MS (pos. mode, CHCA): $[M+Na]^+$ calcd.: 635.4, found: 635.5; $[M+K]^+$ calcd.: 651.4 found: 651.4.

Synthesis of 3,6,9,12,15,18-Hexaoxatriacontanyl α-D-mannopyranosyl-(1→2)-α-D-mannopyranoside (19-2)

Compound **19-2** (27 mg, 35 μmol, 97%) was prepared from **15-2** (38 mg, 36 μmol) according to GP II as a colorless lyophilizate.

TLC: R_f = 0.33 (RP-18-silicagel, MeCN/H_20 3:2); $[\alpha]_D^{20}$ = +32.0 (c = 1.0, MeOH); ^{1}H-NMR (600 MHz, d_4-MeOD): δ = 0.87-0.92 (m, 3H, C*H*$_3$), 1.24-1.39 (m, 18H, 9C*H*$_2$), 1.54-1.60 (m, 2H, 1C*H*$_2$), 3.45-3.49 (m, 2H, OC*H*$_2$$C_{11}H_{23}$), 3.53-3.75 (m, 30H, *H*-3b, *H*-4a, *H*-4b, *H*-5a, *H*-5b, *H*-6a, *H*-6b, 23/2OC*H*$_2$), 3.80-3.86 (m, 5H, *H*-2a, *H*-3a, *H*-6´a, *H*-6´b, 1/2OC*H*$_2$), 3.98 (dd, $^3J_{1b,2b}$ = 1.8 Hz, $^3J_{2b,3b}$ =3.3 Hz, 1H, *H*-2b), 4.96 (d, $^3J_{1b,2b}$ = 1.8 Hz, 1H, *H*-1b), 5.11 (d, $^3J_{1a,2a}$ = 1.7 Hz, 1 H, *H*-1a) ppm; ^{13}C-NMR (151 MHz, d_4-MeOD): δ = 14.50 (1C, *C*H$_3$), 23.79-33.13 (10C, 10*C*H$_2$), 63.13 (1C, *C*-6a or *C*-6b), 63.23 (1C, *C*-6a or *C*-6b), 67.94 (1 C, 1O*C*H_2), 68.88 (1C, *C*-4a or *C*-4b), 69.09 (1C, *C*-4a or *C*-4b), 71.19-71.61 (11C, 11O*C*H_2), 71.93 (1C, *C*-2b) , 72.14 (1C, *C*-3a), 72.43 (1C, *C*-3b or O*C*$H_2C_{11}H_{23}$), 72.48 (1C, *C*-3b or O*C*$H_2C_{11}H_{23}$), 74.72 (1C, *C*-5a or *C*-5b), 75.08 (1C, *C*-5a or *C*-5b), 80.66 (1C, *C*-2a), 100.17 (1C, *C*-1a), 104.28 (1C, *C*-1b) ppm; $C_{36}H_{70}O_{17}$ (774.9); MALDI-MS (pos. mode, CHCA): $[M+Na]^+$ calcd.: 797.5, found: 797.5; $[M+K]^+$ calcd.: 813.5 found: 813.4.

Synthesis of 3,6,9,12,15,18-Hexaoxatriacontanyl α-D-mannopyranosyl-(1→2)- α-D-mannopyranosyl-(1→2)-α-D-mannopyranoside (19-3)

Compound **19-3** (29 mg, 31 μmol, 79%) was prepared from **15-3** (53 mg, 39 μmol) according to GP II as a colorless lyophilizate.

TLC: R_f = 0.60 (RP-18-silicagel, MeCN/$H_2$0 2:1); $[\alpha]_D^{20}$ = +29.1 (c = 1.0, MeOH); ^{1}H-NMR (600 MHz, d_4-MeOD): δ = 0.88-0.92 (m, 3H, C*H_3*), 1.24-1.39 (m, 18H, 9C*H_2*), 1.54-1.60 (m, 2H, 1C*H_2*), 3.45-3.49 (m, 2H, OC*H_2*$C_{11}H_{23}$), 3.52-3.74 (m, 33H, *H*-3c, *H*-4a, *H*-4b, *H*-4c, *H*-5a, *H*-5b, *H*-5c, *H*-6a, *H*-6b, *H*-6c, 23/2OC*H_2*), 3.80-3.88 (m, 5H, *H*-2a, *H*-3a, *H*-3b, *H*-6´a, *H*-6´b, *H*-6´c, 1/2OC*H_2*), 3.97 (dd, $^3J_{1c,2c}$ = 1.7 Hz, $^3J_{2c,3c}$ =3.2 Hz, 1H, *H*-2c), 4.03 (dd, $^3J_{1b,2b}$ = 1.8 Hz, $^3J_{2b,3b}$ =3.1 Hz, 1H, *H*-2b), 4.98 (d, 1H, *H*-1c), 5.10 (d, 1H, *H*-1a), 5.28 (d, 1H, *H*-1b) ppm; ^{13}C-NMR (151 MHz, d_4-MeOD): δ = 14.50 (1C, *C*H_3), 23.78-33.12 (10C, 10*C*H_2), 63.10-63.37 (3C, *C*-6a, *C*-6b, *C*-6c), 67.99 (1 C, 1O*C*H_2), 68.85-69.28 (3C, *C*-4a, *C*-4b, *C*-4c), 71.19-71.60 (11C, 11O*C*H_2), 71.95 (1C, *C*-2c), 72.01 (1C, *C*-3a or *C*-3b), 72.14 (1C, *C*-3a or *C*-3b), 72.43 (1C, *C*-3c or O*C*$H_2C_{11}H_{23}$), 72.51 (1C, *C*-3c or O*C*$H_2C_{11}H_{23}$), 74.67-75.06 (3C, *C*-5a, *C*-5b, *C*-5c), 80.29 (1C, *C*-2b), 80.80 (1C, *C*-2a), 100.13 (1C, *C*-1a), 102.57 (1C, *C*-1b), 104.15 (1C, *C*-1c) ppm; $C_{42}H_{80}O_{22}$ (937.1); MALDI-MS (pos. mode, CHCA): $[M+Na]^+$ calcd.: 959.5, found: 959.9; $[M+K]^+$ calcd.: 975.5 found: 975.9.

Synthesis of 3,6,9,12,15,18-Hexaoxatriacontanyl α-D-mannopyranosyl-(1→2)- α-D-mannopyranosyl-(1→2)- α-D-mannopyranosyl-(1→2)-α-D-mannopyranoside (19-4)

Compound **19-4** (33 mg, 30 μmol, 94%) was prepared from **15-4** (52 mg, 32 μmol) according to GP II as a colorless lyophilizate.

TLC: R_f = 0.33 (RP-18-silicagel, $CH_3CN/H_2$0 3:2); $[\alpha]_D^{20}$ = +30.0 (c = 1.0, MeOH);

TLC: R_f = 0.33 (RP-18-silicagel, $CH_3CN/H_2$0 3:2); $[\alpha]_D^{20}$ = +30.0 (c = 1.0, MeOH); ^{1}H-NMR (600 MHz, d_4-MeOD): δ = 0.88-0.92 (m, 3H, C*H_3*), 1.24-1.38 (m, 18H, 9C*H_2*), 1.54-1.60 (m, 2H, 1C*H_2*), 3.45-3.49 (m, 2H, OC*H_2*$C_{11}H_{23}$), 3.52-3.74 (m, 30H, *H*-3d, *H*-4a, *H*-4b, *H*-4c, *H*-4d, *H*-5a, *H*-5b, *H*-5c, *H*-5d, *H*-6a, *H*-6b, *H*-6c, *H*-6d, 23/2OC*H_2*), 3.79-3.90 (m, 5H, *H*-2a, *H*-3a, *H*-3b, *H*-3c, *H*-6´a, *H*-6´b, *H*-6´c, *H*-6´d, 1/2OC*H_2*), 3.97 (dd, $^3J_{1d,2d}$ = 1.7 Hz, $^3J_{2d,3d}$ =3.3 Hz, 1H, *H*-2d), 4.01-4.04 (m, 2H, *H*-2b, *H*-2c), 4.99 (d, $^3J_{1d,2d}$ = 1.7 Hz, 1H, *H*-1d), 5.10 (d, $^3J_{1a,2a}$ = 1.6 Hz, 1 H, *H*-1a), 5.27 (d, $^3J_{1,2}$ = 1.6 Hz, 1 H, *H*-1b or *H*-1c), 5.28 (d, $^3J_{1,2}$ = 1.6 Hz, 1 H, *H*-1b or *H*-1c) ppm; ^{13}C-NMR (151 MHz, d_4-MeOD): δ = 14.50 (1C, *C*H_3), 23.78-33.12 (10C, 10*C*H_2), 63.08 (1C, *C*-6a or *C*-6b or *C*-6c or *C*-6d), 63.21-63.36 (4C, *C*-6a, *C*-6b, *C*-6c, *C*-6d), 67.99 (1 C, 1O*C*H_2), 68.79-69.25 (4C, *C*-4a, *C*-4b, *C*-4c, *C*-4d), 71.18-71.59 (11C, 11O*C*H_2), 71.95 (1C, *C*-2d) , 72.00-72.17 (1C, *C*-3a, *C*-3b, *C*-3c), 72.43 (1C, *C*-3d or O*C*$H_2C_{11}H_{23}$), 72.51 (1C, *C*-3d or O*C*$H_2C_{11}H_{23}$), 74.67-75.06 (1C, *C*-5a, *C*-5b, *C*-5c, *C*-5d), 80.19 (1C, *C*-2b or *C*-2c), 80.45 (1C, *C*-2b or *C*-2c), 80.79 (1C, *C*-2a), 100.11 (1C, *C*-1a), 102.40 (1C, *C*-1b or *C*-1c), 102.46 (1C, *C*-1b or *C*-1c), 104.11 (1C, *C*-1d) ppm; $C_{48}H_{90}O_{27}$ (1099.2); MALDI-MS (pos. mode, CHCA): $[M+Na]^+$ calcd.: 1121.6, found: 1121.7; $[M+K]^+$ calcd.: 1137.6, found: 1137.7.

Synthesis of 3,6,9,12,15,18-Hexaoxatriacontanyl α-D-mannopyranosyl-(1→2)–α-D-mannopyranosyl-(1→2)–α-D-mannopyranosyl-(1→2)–α-D-mannopyranosyl-(1→2)-α-D-mannopyranoside (19-5)

Compound **19-5** (23 mg, 18 μmol, 86%) was prepared from **15-5** (41 mg, 21 μmol) according to GP II as a colorless lyophilizate.

TLC: R_f = 0.38 (RP-18-silicagel, MeCN/$H_2$0 3:2); $[\alpha]_D^{20}$ = +27.5 (c = 1.0, MeOH);

^{1}H-NMR (600 MHz, $CDCl_3$): δ = 0.88-0.92 (m, 3H, C*H*$_3$), 1.24-1.39 (m, 18H, 9C*H*$_2$), 1.54-1.60 (m, 2H, 1C*H*$_2$), 3.45-3.49 (m, 2H, OC*H*$_2C_{11}H_{23}$), 3.52-3.74 (m, 39H, *H*-3e, *H*-4a, *H*-4b, *H*-4c, *H*-4d, *H*-4e, *H*-5a, *H*-5b, *H*-5c, *H*-5d, *H*-5e, *H*-6a, *H*-6b, *H*-6c, *H*-6d, *H*-6e, 23/2OC*H*$_2$), 3.78-3.90 (m, 11H, *H*-2a, *H*-3a, *H*-3b, *H*-3c, *H*-3d, *H*-6´a, *H*-6´b, *H*-6´c, *H*-6´d, *H*-6´e, 1/2OC*H*$_2$), 3.97 (dd, 1H, *H*-2e), 4.00-4.04 (m, 3H, *H*-2b, *H*-2c, *H*-2d), 4.99 (d, 1H, *H*-1e), 5.10 (d, 1H, *H*-1a), 5.26-5.29 (m, 3H, *H*-1b, *H*-1c, *H*-1d) ppm; ^{13}C-NMR (151 MHz, $CDCl_3$): δ = 14.50 (1C, *C*H_3), 23.79-33.12 (10C, 10*C*H_2), 63.09-63.32 (5C, *C*-6a, *C*-6b, *C*-6c, *C*-6d, *C*-6e), 67.99 (1 C, 1O*C*H_2), 68.80-69.27 (5C, *C*-4a, *C*-4b, *C*-4c, *C*-4d, *C*-4e), 71.20-71.61 (11C, 11O*C*H_2), 71.95 (1C, *C*-2e) , 72.05-72.18 (4C, *C*-3a, *C*-3b, *C*-3c, *C*-3d), 72.43 (1C, *C*-3e or O*C*$H_2C_{11}H_{23}$), 72.51 (1C, *C*-3e or O*C*$H_2C_{11}H_{23}$), 74.68,75.05 (5C, *C*-5a, *C*-5b, *C*-5c, *C*-5d, *C*-5e), 80.18-80.45 (3C, *C*-2b, *C*-2c, *C*-2d), 80.80 (1C, *C*-2a), 100.12 (1C, *C*-1a), 102.32 (1C, *C*-1b or *C*-1c or *C*-1d), 102.41 (1C, *C*-1b or *C*-1c or *C*-1d), 102.45 (1C, *C*-1b or *C*-1c or *C*-1d), 104.13 (1C, *C*-1e) ppm; MALDI-MS (pos. mode, CHCA): $[M+Na]^+$ calcd.: 1283.6, found: 1283.4; $[M+K]^+$ calcd.: 1299.6 found: 1299.4.

Expression and purification of CV-N and its mutant:

The plasmid pET26 harboring the CV-N gene for periplasmic expression was transformed into BL21(DE3) cells. CV-N was expressed on M9 minimal medium containing ^{15}N-NH_4Cl as the sole source of nitrogen supplemented by vitamins. Expression of CV-N was induced after cells reached an OD_{600} (OD=0.7, 600nm) upon addition of isopropyl-1-thio-ß-D-galactopyranoside (1 mM) (IPTG). After induction (22 hr, 22°C) cells were pelleted. The purification took advantage of the extraordinary resistance of CV-N to heat. After resolubilizing the cells the soluble fraction was heated (20min, 65°C) in order to precipitate most other proteins and the remaining soluble supernatant was loaded onto a C18 reverse phase column for purification using a water:acetonitrile gradient in 0.05% TFA. Further purification of CV-N was performed using size exclusion chromatography on a Superdex 75 HR column. Identity of the expressed protein with CV-N was established based on Maldi-TOF MS (exp. mass for ^{15}N-labeled CV-N: 11140.5; calc. mass: 11144.5) and comparison of the [^{15}N,^{1}H]-HSQC spectrum with published data (PDB entry 2ezm)[31].

A mutant of CV-N in which the low-affinity binding site has been removed is presented by the K3N, E23I, N93A mutant of CV-N[35] that was prepared by site-directed mutagenesis using the QuickChange kit (Stratagene;QuickChange TM Site-directed Mutagenesis Kit; La Jolla, CA) according to manufacturer's instructions). Three sets of primers (Microsynth, CH) containing the following sequences were used to introduce the mutations: K3N-forward 5'CTTGGTAACTTCTCCCAGACCTGCTACAACTCCGC3', reverse 5'GCGGAGTTGTAGCAGGTCTGGGAGAAGTTACCAAG3'; E23I- forward 5'CTGACCTCCACCTGCATCCGTACCAACGGTGGTTACAAC3', reverse 5'GTTGTAACCACCGTTGGTACGGATGCAGGTGGAGGTCAG3'; N93A-forward 5'GGACGACCACATCGCTGCCATCGACGGTACCCTGAAATAC3', reverse 5'GTATTTCAGGGTACCGTCGATGGCAGCGATGTGGTCGTCC3'.

The final plasmid encoding these three mutations was transformed into C41 *E. Coli* cells[38], and expressed and purified in an identical manner to that used for wt CV-N. Identity of the mutant CV-N protein was established by Maldi-TOF MS (exp. mass: 11070.9; calc. mass: 11070.5).

NMR spectroscopy:

All NMR spectra were recorded on a BRUKER AV700 spectrometer, equipped with a cryogenically cooled triple resonance probe, at 310K. 2-D [^{15}N,^{1}H]-HSQC experiments[39] were performed with pulsed-field gradients[40] for coherence selection and water suppression and incorporated the sensitivity-enhancement technique described by Palmer[41]. [^{15}N,^{1}H]-HSQC spectra of CV-N when bound to glycolipids **19-2** and **19-3** are essentially identical to those of the protein bound to the corresponding methylglycosides **16-2** and **16-3**, respectively. ^{15}N-^{1}H correlation maps of the complexes were assigned using spectra derived from ^{15}N-edited 3-D NOESY-HSQC and 3-D TOCSY-HSQC experiments[42-44] using the assignments of CV-N as a reference. Similarly, resonances in the mutant CV-N protein were annotated. In particular, sequential $H^{\alpha}_{i}H^{N}_{i+1}$ NOEs were utilized for establishing sequential connectivities. Adding DPC-solution to the dried form of the glycolipid produced the glycolipid-loaded micelles. The latter was prepared from a stock solution in either methanol (**19-2**) or hexafluoroisopropanol (**19-3**). The resulting solution was sonicated (25°C, 15 min). Typically, CV-N samples used for chemical shift mapping experiments contained ^{15}N-uniformly labelled CV-N (0.2-0.3 mM), phosphate buffer (20 mM, pH 6.5), and d_{38}-dodecylphosphocholine (300 mM) (DPC). Samples used for assignment purposes of the ^{15}N-^{1}H correlation maps contained CV-N (0.5-1.0 mM). A table including the resonance assignments for wild type and mutant-CV-N is provided in the Supp. Mat. Spectra were processed using TOPSPIN and then transferred into the programs XEASY[45] or SPARKY[46] in the case of 2-D spectra and into CARA[47] in the case of 3-D spectra for further analysis.

Integration of the glycolipids into the micelles was verified from a reduction of the translational diffusion constant in the mixed micelles, as measured by a BPP-LED sequence[48] modified to reduce effects due to convection[49]. In those measurements the diffusion delay was set to 200 ms and the gradients (0.05-40 G/cm) were applied as sine shapes. Gradient strengths were calibrated using a phantom sample. Internal software routines in TOPSPIN were used to derive the corresponding diffusion coefficients.

Acknowledgements

We greatly acknowledge financial support by the Swiss National Fond (grant 3100A0-100462 to J.M. and A.N.), the Intramural AIDS Targeted Antiviral Program, NIH (C.A.B) and the Deutsche Forschungsgemeinschaft (J.U.M.). We would like to thank K. Zerbe for her continuous support during the molecular biology part of this project.

References:

1. Rudd, P. M., Wormald, M. R. & Dwek, R. A. (2004) Sugar-mediated ligand-receptor interactions in the immune system, *Trends. Biotechnol. 22*, 524-30.
2. Sharon, N. & Lis, H. (1989) Lectins as cell recognition molecules, *Science. 246*, 227-34.
3. Lis, H. & Sharon, N. (1998) Lectins: Carbohydrate-Specific Proteins That Mediate Cellular Recognition, *Chem. Rev. 98*, 637-674.
4. Brandley, B. K. (1991) Cell surface carbohydrates in cell adhesion, *Semin. Cell Biol. 2*, 281-7.
5. Bax, A. (1989) Two-Dimensional NMR and Protein Structure, *Annu. Rev. Biochem. 58*, 223-256.
6. Wüthrich, K. (2003) NMR studies of structure and function of biological macromolecules, *Angew. Chem. Int. Ed. Engl. 42*, 3340-63.
7. Ferentz, A. E. & Wagner, G. (2000) NMR spectroscopy: a multifaceted approach to macromolecular structure, *Q. Rev. Biophys. 33*, 29-65.
8. Wüthrich, K. (1986) *NMR of Proteins and Nucleic Acids*, 1st edn, Wiley, New York.
9. Damberg, P., Jarvet, J. & Gräslund, A. (2001) Micellar systems as solvents in peptide and protein structure determination, *Methods Enzymol. 339*, 271-85.
10. Baleja, J. D. (2001) Structure determination of membrane-associated proteins from nuclear magnetic resonance data, *Anal. Biochem. 288*, 1-15.
11. Bader, R., Lerch, M. & Zerbe, O. (2002) NMR of membrane-associated peptides and proteins in *BioNMR in Drug Research* (Zerbe, O., ed) pp. 95-120, Wiley-VCH, Weinheim.
12. Opella, S. J. (1997) NMR and membrane proteins, *Nature Struct. Biol. 4*, 845-8.
13. Opella, S. J., Kim, Y. & McDonnell, P. (1994) Experimental nuclear magnetic resonance studies of membrane proteins, *Methods Enzymol. 239*, 536-60.
14. Tugarinov, V., Hwang, P. M. & Kay, L. E. (2004) Nuclear magnetic resonance spectroscopy of high-molecular-weight proteins, *Annu. Rev. Biochem. 73*, 107-46.
15. Wider, G. (2005) NMR techniques used with very large biological macromolecules in solution, *Methods Enzymol. 394*, 382-98.
16. Sanders, C. R. & Oxenoid, K. (2000) Customizing model membranes and samples for NMR spectroscopic studies of complex membrane proteins, *Biochim. Biophys. Acta. 1508*, 129-45.
17. Vinogradova, O., Sonnichsen, F. & Sanders, C. R., 2nd. (1998) On choosing a detergent for solution NMR studies of membrane proteins, *J. Biomol. NMR. 11*, 381-6.
18. Krüger-Koplin, R. D., Sorgen, P. L., Krüger-Koplin, S. T., Rivera-Torres, I. O., Cahill, S. M., Hicks, D. B., Grinius, L., Krulwich, T. A. & Girvin, M. E. (2004) An evaluation of detergents for NMR structural studies of membrane proteins, *J. Biomol. NMR. 28*, 43-57.
19. Lauterwein, J., Bösch, C., Brown, L. R. & Wüthrich, K. (1979) Physicochemical studies of the protein-lipid interactions in melittin-containing micelles, *Biochim. Biophys. Acta. 556*, 244-264.
20. Taylor, M. E. & Drickamer, K. (2003) *Introduction to glycobiology*, Oxford University Press, Oxford.

21. Boyd, M. R., Gustafson, K. R., McMahon, J. B., Shoemaker, R. H., O'Keefe, B. R., Mori, T., Gulakowski, R. J., Wu, L., Rivera, M. I., Laurencot, C. M., Currens, M. J., Cardellina, J. H., 2nd, Buckheit, R. W., Jr., Nara, P. L., Pannell, L. K., Sowder, R. C., 2nd & Henderson, L. E. (1997) Discovery of cyanovirin-N, a novel human immunodeficiency virus-inactivating protein that binds viral surface envelope glycoprotein gp120: potential applications to microbicide development, *Antimicrob. Agents Chemother. 41*, 1521-30.

22. Mayer, T. K., Kratzer, B. & Schmidt, R. R. (1994) Synthesis of a GPI Anchor of Yeast (Saccharomyces cerevisiae), *Angew. Chem. Int. Ed. Engl. 33*, 2177-2181.

23. Yamazuki, F., Sato, S., Nukada, T. & Ito, Y. (1990) Synthesis of α-D-Manp-(1→3)-[β-D-GlcpNAc-(1→4)]-[α-D-Manp-(1→6)]-β-D-Manp-(1→4)-β-D-GlcpNAc-(1→4)-[α-L-Fucp-(1→6)]-D-GlcpNAc, a core glycoheptaose of a "bisected" complex-type glycan of glycoproteins, *Carbohydr. Res. 201*, 31-50.

24. Rademann, J. & Schmidt, R. R. (1997) Repetitive Solid Phase Glycosylation on an Alkyl Thiol Polymer Leading to Sugar Oligomers Containing 1-2-trans- and 1-2-cis-Glycosidic Linkages, *J. Org. Chem. 62*, 3650-3653.

25. Gao, X. & Wong, T. C. (1998) Studies of the binding and structure of adrenocorticotropin peptides in membrane mimics by NMR spectroscopy and pulsed-field gradient diffusion, *Biophys. J. 74*, 1871-88.

26. Eckert, D. M. & Kim, P. S. (2001) Mechanisms of viral membrane fusion and its inhibition, *Annu. Rev. Biochem. 70*, 777-810.

27. Chan, D. C. & Kim, P. S. (1998) HIV entry and its inhibition, *Cell. 93*, 681-4.

28. Blumenthal, R., Clague, M. J., Durell, S. R. & Epand, R. M. (2003) Membrane fusion, *Chem. Rev. 103*, 53-69.

29. Botos, I., O'Keefe, B. R., Shenoy, S. R., Cartner, L. K., Ratner, D. M., Seeberger, P. H., Boyd, M. R. & Wlodawer, A. (2002) Structures of the complexes of a potent anti-HIV protein cyanovirin-N and high mannose oligosaccharides, *J. Biol. Chem. 277*, 34336-42.

30. Bewley, C. A. (2001) Solution structure of a cyanovirin-N:Man alpha 1-2Man alpha complex: structural basis for high-affinity carbohydrate-mediated binding to gp120, *Structure. 9*, 931-40.

31. Bewley, C. A., Gustafson, K. R., Boyd, M. R., Covell, D. G., Bax, A., Clore, G. M. & Gronenborn, A. M. (1998) Solution structure of cyanovirin-N, a potent HIV-inactivating protein, *Nat. Struct. Biol. 5*, 571-8.

32. Bewley, C. A. & Otero-Quintero, S. (2001) The potent anti-HIV protein cyanovirin-N contains two novel carbohydrate binding sites that selectively bind to Man(8) D1D3 and Man(9) with nanomolar affinity: implications for binding to the HIV envelope protein gp120, *J. Am. Chem. Soc. 123*, 3892-902.

33. Bewley, C. A., Kiyonaka, S. & Hamachi, I. (2002) Site-specific discrimination by cyanovirin-N for alpha-linked trisaccharides comprising the three arms of Man(8) and Man(9), *J. Mol. Biol. 322*, 881-9.

34. Sandstrom, C., Berteau, O., Gemma, E., Oscarson, S., Kenne, L. & Gronenborn, A. M. (2004) Atomic mapping of the interactions between the antiviral agent cyanovirin-N and oligomannosides by saturation-transfer difference NMR, *Biochemistry. 43*, 13926-31.

35. Chang, L. C. & Bewley, C. A. (2002) Potent inhibition of HIV-1 fusion by cyanovirin-N requires only a single high affinity carbohydrate binding site: characterization of low affinity carbohydrate binding site knockout mutants, *J. Mol. Biol. 318*, 1-8.

36. Wimley, W. C. & White, S. H. (1996) Experimentally determined hydrophobicity scale for proteins at membrane interfaces, *Nature Struct. Biol. 3*, 842-848.

37. Killian, J. A. & von Heijne, G. (2000) How proteins adapt to a membrane-water interface, *TIBS. 25*, 429-434.

38. Miroux, B. & Walker, J. E. (1996) Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels, *J. Mol. Biol. 260*, 289-98.

39. Bodenhausen, G. & Ruben, D. J. (1980) Natural Abundance Nitrogen-15 NMR by Enhanced Heteronuclear Spectroscopy, *Chem. Phys. Lett. 69*, 185-189.

40. Keeler, J., Clowes, R. T., Davis, A. L. & Laue, E. D. (1994) Pulsed-field gradients: theory and practice, *Methods Enzymol. 239*, 145-207.

41. Palmer, A. G., Cavanagh, J., Wright, P. E. & Rance, M. (1991) Sensitivity Improvement in Proton-Detected Two-Dimensional Heteronuclear Correlation NMR Spectroscopy, *J. Mag. Reson. 93*, 151-170.

42. Sattler, M., Schleucher, J. & Griesinger, C. (1999) Heteronuclear multidimensional NMR experiments for the structure determination of proteins in solution employing pulsed field gradients, *Prog. NMR Spectrosc. 34*, 93-158.

43. Marion, D., Kay, L. E. & Sparks, S. W. (1989) Three-Dimensional Heteronuclear NMR of ^{15}N-labelled Proteins, *J. Am. Chem. Soc. 111*, 1515-1517.

44. Marion, D., Driscoll, P. C., Kay, L. E., Wingfield, P. T., Bax, A., Gronenborn, A. M. & Clore, G. M. (1989) Overcoming the Overlap Problem in the Assignment of ^{1}H NMR Spectra of Larger Proteins by Use of Three-Dimensional Heteronuclear ^{1}H-^{15}N Hartmann-Hahn Multiple-Quantum Coherence and Nuclear Overhauser-Multiple Quantum Coherence Spectroscopy: Application to Interleukin 1β, *Biochemistry. 28*, 6150-6156.

45. Bartels, C., Xia, T.-h., Billeter, M., Güntert, P. & Wüthrich, K. (1995) The program XEASY for computer-supported spectral analysis of biological macromolecules, *J. Biomol. NMR. 6*, 1-10.

46. Goddard, T. D. & Kneller, D. G. SPARKY 3 in, University of California, San Francisco

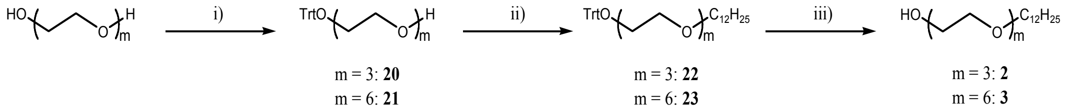
47. Keller, R. (2004) *The Computer Aided Resonance Assignment*, CANTINA Verlag.

48. Wu, D. H., Chen, A. D. & Johnson, C. S. (1995) An Improved Diffusion-Ordered Spectroscopy Experiment Incorporating Bipolar-Gradient Pulses, *J. Magn. Reson. Ser. A. 115*, 260-264.

49. Jerschow, A. & Müller, N. (1997) Suppression of convection artifacts in stimulated-echo diffusion experiments. Double-stimulated-echo experiments, *J. Magn. Reson. 125*, 372-375.

7.4 Supplementary Material

Experimental details for compounds 2, 3, 4-1, 4-2, 4-3, 5-1, 5-2, 6-1, 6-2, 7-1, 7-2, 7-3, 7-4, 7-5, 12-2, 12-3, 13-2, 14-2, 15-1, 15-2, 15-3, 15-4, 15-5, 15-6, 16-2, 16-3, 17-2, 18-2, 19-1, 19-2, 19-3, 19-4, 19-5, 20, 21, 22 and 23:



Scheme outlining the synthetic route to the linker moieties. Reagents used are i) TrtCl, DMAP, pyr; ii) $C_{12}H_{25}Br$, NaH, TBAI, THF; iii) AcOH/MeOH.

3,6,9-Trioxahenicosanol (2)[1,2]

3,6,9,12,15-Hexaoxatriacontanol (3)[2]

TLC: R_f = 0.60 (petroleum ether/ethyl acetate 4:1); $[\alpha]_D^{20}$ = +19.7 (c = 1.0, $CHCl_3$); ^{1}H-NMR (600 MHz, $CDCl_3$): δ = 0.88 (pseudo t, 3H, C$\mathit{H_3}$), 1.20-1.32 (m, 18H, 9C$\mathit{H_2}$), 1.55 (pseudo quint, 2H, C$\mathit{H_2}$), 2.15 (s, 3H, COC$\mathit{H_3}$), 3.37-3.42 (m, 1H, 1/2O$\mathit{C}H_2C_{11}H_{23}$), 3.63-3.72 (m, 2H, *H*-6, 1/2O$\mathit{C}H_2C_{11}H_{23}$), 3.78-3.83 (m, 2H, *H*-5, *H*-6′), 3.88 (dd, 1H, *H*-4), 3.99 (dd, $^3J_{2,3}$ = 3.4 Hz, $^3J_{3,4}$ = 9.3 Hz, 1H, *H*-3), 4.46-4.73 (m, 4H, 2PhC$\mathit{H_2}$), 4.83 (d, $^3J_{1,2}$ = 1.8 Hz, 1 H, *H*-1), 4.86 (d, 1H, 1PhC$\mathit{H_2}$), 5.36 (dd, $^3J_{1,2}$ = 1.8 Hz, $^3J_{2,3}$ = 3.4 Hz, 1H, *H*-2), 7.15-7.37 (m, 9H, *Ph*) ppm; ^{13}C-NMR (151 MHz, $CDCl_3$): δ = 14.11 (1C, *C*H_3), 21.15 (1C, CO*C*H_3), 22.68-31.97 (10C, 10*C*H_2), 68.01 (1C, O*C*$H_2C_{11}H_{23}$), 68.89 (2C, *C*-2, *C*-6), 71.28 (1C, *C*-5), 71.78-73.41 (2C, 2*C*H_2Ph), 74.39 (1C, *C*-4), 75.20 (1C, 1*C*H_2Ph), 78.30 (1C, *C*-3), 97.71 (1C, *C*-1), 127.54-138.36 (18C, *Ph*), 170.55 (1C, *C*OCH_3), ppm; elemental analysis is calcd. (%) for $C_{41}H_{56}O_7$ (660.9): C: 74.51, H: 8.54; found: C: 74.20, H: 8.33; $C_{41}H_{56}O_7$ (660.9); MALDI-MS (pos. mode, CHCA): $[M+Na]^+$ calcd.: 683.4, found: 683.8; $[M+K]^+$ calcd.: 699.4, found: 699.8.

Compound **22** (7.45 g, 13.3 mmol) was dissolved in 150 ml methanol/acetic acid (2:1 v/v) and the solution was refluxed for 16h. After cooling to RT the reaction mixture was concentrated under reduced pressure and co-evaporated with toluene. The residue was purified by flash column chromatography (silica gel, toluene/ethyl acetate 3:1 → 1:1) to afford compound **2** (3.86 g, 12.1 mmol, 91%).

Compound **23** (14.3 g, 20.6 mmol) was dissolved in 300 ml methanol/acetic acid (2:1 v/v) and the solution was refluxed for 16h. After cooling to RT the reaction mixture was concentrated under reduced pressure and co-evaporated with toluene. The residue was purified by flash column chromatography (silica gel, toluene/acetone 2:1 → 1:1 → 1:2) to afford compound **3** (8.7 g, 19.3 mmol, 94%).

Synthesis of Methyl 2-*O*-acetyl-3,4,6-tri-*O*-benzyl-α-D-mannopyranoside (4-1)[3]

Compound **4-1** was prepared from methanol (20 ml, 494 μmol) and the mannosyl donor **1** (3.64 g, 5.71 mmol) according to GP I. The residue was purified by flash column chromatography (silica gel, petroleum ether/ethyl acetate 8:1 → 5:1) to afford **4-1** (2.54 g, 5.01 mmol, 88%) as a pale, colorless oil.

Synthesis of Methyl 2-*O*-acetyl-3,4,6-tri-*O*-benzyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-benzyl-α-D-mannopyranoside (4-2)[4,5]

Compound **4-1** (1.78 g, 3.51 mmol) was deprotected according to GP II to afford **8-1** (1.53 g, 3.29 mmol, 94%).

Compound **4-2** was prepared from **8-1** (1.53 g, 3.29 mmol) and the mannosyl donor **1** (2.73 g, 4.29 mmol) according to GP I. The residue was purified by flash column chromatography (silica gel, petroleum ether/ethyl acetate 5:1 → 3:1) to afford **4-2** (2.83 g, 3.01 mmol, 91%) as a pale, colorless oil.

Synthesis of Methyl 2-*O*-acetyl-3,4,6-tri-*O*-benzyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-benzyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-benzyl-α-D-mannopyranoside (4-3)[6]

Compound **4-2** (1.80 g, 1.92 mmol) was deprotected according to GP II to afford **8-2** (1.59 g, 1.77 mmol, 92%).

Compound **4-3** was prepared from **8-2** (1.59 g, 1.77 mmol) and the mannosyl donor **1** (1.47 g, 2.31 mmol) according to GP I. The residue was purified by flash column chromatography (silica gel, petroleum ether/ethyl acetate 8:1 → 3:1) to afford **4-3** (1.84 g, 1.34 mmol, 76%) as pale, colorless oil in sufficient purity for further conversions. For analytical data a small portion was purified by MPLC (silica gel, petroleum ether/ethyl acetate 4:1).

Synthesis of Dodecyl 2-*O*-acetyl-3,4,6-tri-*O*-benzyl-α-D-mannopyranoside (5-1)

Compound **5-1** was prepared from dodecanol (1.10 g, 5.90 mmol) and the mannosyl donor **1** (4.45 g, 6.99 mmol) according to GP I. The residue was purified by flash column chromatography (silica gel, petroleum ether/ethyl acetate 10:1) to afford **5-1** (3.41 g, 5.16 mmol, 87%) as a pale, colorless oil.

Synthesis of Dodecyl 2-*O*-acetyl-3,4,6-tri-*O*-benzyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-benzyl-α-D-mannopyranoside (5-2)

Compound **5-1** (3.40 g, 5.14 mmol) was deprotected according to GP II to afford **9-1** (3.01 g, 4.86 mmol, 96%).

Compound **5-2** was prepared from **9-1** (780 mg, 1.26 mmol) and the mannosyl donor **1** (1.04 g, 1.63 mmol) according to GP I. The residue was purified by flash column chromatography (silica gel, petroleum ether/ethyl acetate 8:1 → 5:1) to afford **5-2** (1.08 g, 988 μmol, 78%) as a pale, colorless oil.

TLC: R_f = 0.78 (petroleum ether/ethyl acetate 2:1); $[\alpha]_D^{20}$ = +18.6 (c = 1.0, $CHCl_3$); ^{1}H-NMR (600 MHz, $CDCl_3$): δ = 0.89 (pseudo t, 3H, CH_3), 1.20-1.32 (m, 18H, 9CH_2), 1.50 (pseudo quint, 2H, 1CH_2), 2.13 (s, 3H, $COCH_3$), 3.25-3.30 (m, 1H, 1/2O$CH_2C_{11}H_{23}$), 3.57-3.62 (m, 1H, 1/2O$CH_2C_{11}H_{23}$), 3.69-3.81 (m, 5H, *H*-5a or *H*-5b, *H*-6a, *H*-6a´, *H*-6b, *H*-6´b), 3.82-3.87 (m, 2H, *H*-4a, *H*-4b), 3.91-3.94 (dd, $^3J_{2a,3a}$ = 3.0 Hz, $^3J_{3a,4a}$ = 9.6 Hz, 1H, *H*-3a), 3.96-4.01 (m, 3H, *H*-2a, *H*-3b, *H*-5a or *H*-5b), 4.39-4.70 (m, 10H, 5PhCH_2), 4.84-4.88 (m, 3H, *H*-1a, 1PhCH_2), 5.10 (d, 1H, *H*-1b), 5.55 (dd, $^3J_{1b,2b}$ = 1.9 Hz, $^3J_{2b,3b}$ = 3.3 Hz, 1H, *H*-2b), 7.15-7.37 (m, 30H, *Ph*) ppm; ^{13}C-NMR (151 MHz, $CDCl_3$): δ = 14.11 (1C, CH_3), 21.12 (1C, $COCH_3$), 22.67-31.91 (10C, 10CH_2), 67.72 (1C, O$CH_2C_{11}H_{23}$), 68.72 (1C, *C*-2b), 69.03 (1C, *C*-6a or *C*-6b), 69.28 (1C, *C*-6a or *C*-6b), 71.73 (1C, *C*-5a or *C*-5b), 71.75 (1C, *C*-5a or *C*-5b), 71.88-73.35 (4C, 4CH_2Ph), 74.35 (1C, *C*-4a or *C*-4b), 74.68 (1C, *C*-4a or *C*-4b), 74.94 (1C, *C*-2a), 75.01-75.15 (2C, 2CH_2Ph), 78.13 (1C, *C*-3b), 79.77 (1C, *C*-3a), 98.61 (1C, *C*-1a), 99.52 (1C, *C*-1b), 127.34-138.53 (36C, *Ph*), 170.93 (1C, $COCH_3$) ppm; ; elemental analysis is calcd. (%) for $C_{68}H_{84}O_{12}$ (1093.4): C: 74.70, H: 7.74; found: C: 74.66, H: 8.02; MALDI-MS (pos. mode, CHCA): $[M+Na]^+$ calcd.: 1115.6, found: 1115.3; $[M+K]^+$ calcd.: 1131.7, found: 1131.4.

Synthesis of 3,6,9-Trioxahenicosanyl 2-*O*-acetyl-3,4,6-tri-*O*-benzyl-α-D-mannopyranoside (6-1)

Compound **6-1** was prepared from the glycol **2** (1.00 g, 3.14 mmol) and the mannosyl donor **1** (2.39 g, 3.75 mmol) according to GP I. The residue was purified by flash column chromatography (silica gel, toluene/acetone 20:1 → 10:1) to afford **6-1** (2.34 g, 2.95 mmol, 94%) as a pale, colorless oil.

Synthesis of 3,6,9-Trioxahenicosanyl 2-*O*-acetyl-3,4,6-tri-*O*-benzyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-benzyl-α-D-mannopyranoside (6-2)

Compound **6-1** (2.18 g, 2.75 mmol) was deprotected according to GP II to afford **10-1** (2.00 g, 2.66 mmol, 97%).

Compound **6-2** was prepared from **10-1** (1.16 g, 1.54 mmol) and the mannosyl donor **1** (1.18 g, 1.85 mmol) according to GP I. The residue was purified by flash column chromatography (silica gel, toluene/acetone 20:1 → 10:1) to afford **6-2** (1.14 g, 930 μmol, 60%) as a pale, colorless oil.

Synthesis of 3,6,9,12,15,18-Hexaoxatriacontanyl 2-*O*-acetyl-3,4,6-tri-*O*-benzyl-α-D-mannopyranoside (7-1)

Compound **7-1** was prepared from the glycol **3** (1.43 g, 3.17 mmol) and the mannosyl donor **1** (2.46 g, 3.86 mmol) according to GP I. The residue was purified by flash column chromatography (silica gel, toluene/acetone 5:1 → 3:1) to afford **7-1** (2.56 g, 2.77 mmol, 87%) as a pale, colorless oil.

Synthesis of 3,6,9,12,15,18-Hexaoxatriacontanyl 2-*O*-acetyl-3,4,6-tri-*O*-benzyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-benzyl-α-D-mannopyranoside (7-2)

Compound **7-1** (2.24 g, 2.42 mmol) was deprotected according to GP II to afford **11-1** (1.92 g, 2.17 mmol, 90%).

Compound **7-2** was prepared from **11-1** (1.60 g, 1.81 mmol) and the mannosyl donor **1** (1.51 g, 2.37 mmol) according to GP I. The residue was purified by flash column chromatography (silica gel, toluene/acetone 8:1) to afford **7-2** (1.66 g, 1.22 mmol, 67%) as a pale,colorless oil in sufficient purity for further conversions. For analytical data a small portion was purified by MPLC (silica gel, toluene/acetone 8:1).

Synthesis of 3,6,9,12,15,18-Hexaoxatriacontanyl 2-*O*-acetyl-3,4,6-tri-*O*-benzyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-benzyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-benzyl-α-D-mannopyranoside (7-3)

Compound **7-2** (1.35 g, 995 μmol) was deprotected according to GP II to afford **11-2** (1.18 g, 897 mmol, 90%).

Compound **7-3** was prepared from **11-2** (1.07 g, 813 μmol) and the mannosyl donor **1** (622 mg, 977 μmol) according to GP I. The residue was purified by flash column chromatography (silica gel, toluene/acetone 10:1 → 5:1) to afford **7-3** (1.09 g, 609 μmol, 75%) as pale, colorless oil in sufficient purity for further conversions. For analytical data a small portion was purified by MPLC (silica gel, toluene/acetone 8:1).

Synthesis of 3,6,9,12,15,18-Hexaoxatriacontanyl 2-*O*-acetyl-3,4,6-tri-*O*-benzyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-benzyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-benzyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-benzyl-α-D-mannopyranoside (7-4)

Compound **7-3** (795 mg, 444 μmol) was deprotected according to GP II to afford **11-3** (712 mg, 407 μmol, 92%).

Compound **7-4** was prepared from **11-3** (564 mg, 323 μmol) and the mannosyl donor **1** (250 mg, 392 μmol) according to GP I. The residue was purified by flash column chromatography (silica gel, toluene/acetone 10:1 → 8:1) to afford **7-4** (552 mg, 248 μmol, 77%) as pale colorless oil in sufficient purity for further conversions. For analytical data a small portion was purified by MPLC (silica gel, toluene/acetone 8:1).

TLC: $R_f = 0.48$ (toluene/acetone 5:1); $[\alpha]_D^{20} = +15.8$ ($c = 1.0$, $CHCl_3$); ^{1}H-NMR (600 MHz, $CDCl_3$): $\delta = 0.89$ (pseudo t, 3H, $C\mathit{H_3}$), 1.20-1.36 (m, 18H, 9$C\mathit{H_2}$), 1.58 (pseudo quint, 2H, 1$C\mathit{H_2}$), 2.14 (s, 3H, $COC\mathit{H_3}$), 3.41-3.80 (m, 38H, *H*-4a, *H*-4b, *H*-4c, *H*-5a, *H*-6a, *H*-6b, *H*-6c, *H*-6d, *H*-6´a, *H*-6´b, *H*-6´c, *H*-6´d, 12$OC\mathit{H}_2$, $OC\mathit{H}_2C_{11}H_{23}$), 3.83-3.98 (m, 7H, *H*-3a, *H*-3b, *H*-3c, *H*-4d, *H*-5b, *H*-5c, *H*-5d), 4.00-4.04 (m, 2H, *H*-2a, *H*-3d), 4.08-4.18 (m, 3H, *H*-2b, *H*-2c, 1/2$PhC\mathit{H_2}$), 4.31-4.88 (m, 23H, 23/2$PhC\mathit{H_2}$), 4.96 (d, 1H, *H*-1a), 5.05 (d, 1H, *H*-1d), 5.18 (d, 1H, *H*-1c), 5.23 (d, 1H, *H*-1b), 5.57 (dd, $^3J_{1d,2d} = 1.8$ Hz, $^3J_{2d,3d} = 3.0$ Hz, 1H, *H*-2d), 7.03-7.35 (m, 60H, *Ph*) ppm; ^{13}C-NMR (151 MHz, $CDCl_3$): $\delta = 14.11$ (1C, $\mathit{C}H_3$), 21.15 (1C, $CO\mathit{C}H_3$), 22.66-31.88 (10C, 10$\mathit{C}H_2$), 66.62 (1C, 1$O\mathit{C}H_2$), 68.57-69.50 (5C, *C*-2e, *C*-6a, *C*-6b, *C*-6c, *C*-6d), 69.99-70.56 (14C, 11$O\mathit{C}H_2$), 71.51 (1C, $O\mathit{C}H_2C_{11}H_{23}$), 71.60-72.18 (8C, *C*-5a, *C*-5b, *C*-5c, *C*-5d, 4$\mathit{C}H_2Ph$), 73.19-73.29 (4C, 4$\mathit{C}H_2Ph$), 74.18 (1C, *C*-4d), 74.58-75.14 (8C, *C*-2c, *C*-4a, *C*-4b, *C*-4c, 4$\mathit{C}H_2Ph$), 75.42 (1C, *C*-2b), 75.50 (1C, *C*-2a), 78.23 (1C, *C*-3d), 79.21 (2C, *C*-3a, *C*-3b), 79.31 (1C, *C*-3c), 98.79 (1C, *C*-1a), 99.33 (1C, *C*-1d), 100.68 (1C, *C*-1c), 101.09 (1C, *C*-1b), 127.28-138.55 (60C, *Ph*), 170.08 (1C, $\mathit{C}OCH_3$) ppm; elemental analysis is calcd. (%) for $C_{134}H_{164}O_{28}$ (2222.7): C: 72.41, H: 7.44; found: C: 72.43, H: 7.42; MALDI-MS (pos. mode, CHCA): $[M+Na]^+$ calcd.: 2245.7, found: 2245.4; $[M+K]^+$ calcd.: 2261.8, found: 2261.2.

Synthesis of 3,6,9,12,15,18-Hexaoxatriacontanyl 2-*O*-acetyl-3,4,6-tri-*O*-benzyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-benzyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-benzyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-benzyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-benzyl-α-D-mannopyranoside (7-5)

Compound **7-4** (307 mg, 138 μmol) was deprotected according to GP II to afford **11-4** (259 mg, 119 μmol, 86%).

Compound **7-5** was prepared from **11-4** (259 mg, 119 μmol) and the mannosyl donor **1** (100 mg, 157 mmol) according to GP I. The residue was purified by flash column chromatography (silica gel, toluene/acetone 10:1 → 8:1) to afford **7-5** (235 mg, 89 μmol, 75%) as pale, colorless oil in sufficient purity for further conversions. For analytical data a small portion was purified by MPLC (silica gel, toluene/acetone 8:1).

Synthesis of Methyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-acetyl-α-D-mannopyranoside (12-2)[7]

Compound **12-2** was prepared from **4-2** (1.06 g, 1.13 mmol) according to GP III. The residue was purified by flash column chromatography (silica gel, petroleum ether/ethyl acetate 3:1 → 1:1) to afford **12-2** (449 mg, 690 μmol, 61%) as a foam.

Synthesis of Methyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-acetyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-acetyl-α-D-mannopyranoside (12-3)

Compound **12-3** was prepared from **4-3** (380 mg, 277 μmol) according to GP III. The residue was purified by flash column chromatography (silica gel, petroleum ether/ethyl acetate 1:1 → 1:2), followed by MPLC (silica gel, toluene/acetone 5:2) to afford **12-3** (116 mg, 124 μmol, 45%) as a foam.

Synthesis of Dodecyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-acetyl-α-D-mannopyranoside (13-2)

Compound **13-2** was prepared from **5-2** (630 mg, 576 μmol) according to GP III. The residue was purified by flash column chromatography (silica gel, toluene/acetone 5:1 → 3:1) to afford **13-2** (413 mg, 513 μmol, 89%) as a pale, colorless oil.

Synthesis of 3,6,9-Trioxahenicosanyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-acetyl-α-D-mannopyranoside (14-2)

Compound **14-2** was prepared from **6-2** (1.45 g, 1.18 mmol) according to GP III. The residue was purified by flash column chromatography (silica gel, toluene/acetone 5:1 → 3:1) to afford 770 mg of a mixture enriched in 14-2. A portion (240 mg) was subjected to further purification by MPLC (silica gel, toluene/acetone 5:1) to afford **14-2** (132 mg, 141 μmol, 38%) as a foam.

TLC: R_f = 0.47 (toluene/acetone 3:1); $[\alpha]_D^{20}$ = +38.7 (c = 1.0, $CHCl_3$); ^{1}H-NMR (600 MHz, $CDCl_3$): δ = 0.86 (pseudo t, 3H, C*H*$_3$), 1.19-1.33 (m, 18H, 9C*H*$_2$), 1.55 (pseudo quint, 2H, 1C*H*$_2$), 1.99-2.14 (7s, 21H, 7COC*H*$_3$), 3.42 (pseudo t, 2H, OC*H*$_2$C$_{11}$H$_{23}$), 3.54-3.57 (m, 2H, OC*H*$_2$), 3.60-3.66 (m, 9H, 9/2OC*H*$_2$), 3.79-3.82 (m, 1H, 1/2OC*H*$_2$), 3.98 (ddd, $^3J_{4a,5a}$ = 9.4 Hz, $^3J_{5a,6a}$ = 3.7 Hz, $^3J_{5a,6'a}$ = 2.5 Hz, 1H, *H*-5a), 4.04 (dd, 1H, *H*-2a), 4.07-4.12 (m, 2H, *H*-6a, *H*-6b), 4.15 (ddd, $^3J_{4b,5b}$ = 10.0 Hz, $^3J_{5b,6b}$ = 5.4 Hz, $^3J_{5b,6'b}$ = 2.1 Hz, 1H, *H*-5b), 4.19-4.23 (m, 2H, *H*-6´a, *H*-6´b), 4.91 (d, $^3J_{1b,2b}$ = 1.3 Hz, 1H, *H*-1b), 4.96 (d, $^3J_{1a,2a}$ = 1.3 Hz, 1H, *H*-1a), 5.24-

5.35 (m, 4H, *H*-2b, *H*-3a, *H*-4a, *H*-4b), 5.39 (dd, $^3J_{2b,3b}$ = 3.4 Hz, $^3J_{3b,4b}$ = 10.0 Hz, 1H, *H*-3b) ppm; ^{13}C-NMR (151 MHz, $CDCl_3$): δ = 14.07 (1C, CH_3), 20.62-20.84 (7C, 7COCH_3), 22.63-31.85 (10C, 10CH_2), 62.02 (1C, *C*-6a or *C*-6b), 62.45 (1C, *C*-6a or *C*-6b), 66.09 (1C, *C*-4a), 66.32 (1C, *C*-4b), 67.36(1C, 1OCH_2), 68.32 (1C, *C*-3b), 68.39 (1C, *C*-5a), 69.06 (1C, *C*-5b), 69.72 (1C, *C*-2b), 69.92-69.97 (2C, 2OCH_2), 70.22 (1C, *C*-3a), 70.54-70.66 (3C, 3OCH_2), 71.50 (1C, O$CH_2C_{11}H_{23}$), 77.04 (1C, *C*-2a), 98.30 (1C, *C*-1a), 99.12 (1C, *C*-1b), 169.32-170.86 (7C, 7*C*OCH_3) ppm; $C_{44}H_{72}O_{21}$ (937.0); elemental analysis is calcd. (%) for $C_{44}H_{72}O_{21}$ (937.0): C: 56.40, H: 7.74; found: C: 56.44, H: 7.58; MALDI-MS (pos. mode, CHCA): $[M+Na]^+$ calcd.: 959.5, found: 959.0; $[M+K]^+$ calcd.: 975.5, found: 975.0.

Synthesis of 3,6,9,12,15,18-Hexaoxatriacontanyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranoside (15-1)

Compound **15-1** was prepared from **7-1** (380 mg, 411 μmol) according to GP III. The residue was purified by flash column chromatography (silica gel, toluene/acetone 5:2) to afford **15-1** (196 mg, 251 μmol, 61%) as a pale, colourless oil.

Synthesis of 3,6,9,12,15,18-Hexaoxatriacontanyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-acetyl-α-D-mannopyranoside (15-2)

Compound **15-2** was prepared from **7-2** (817 mg, 602 μmol) according to GP III. The residue was purified by flash column chromatography (silica gel, toluene/acetone 5:2 → 2:1) followed by MPLC (silica gel, toluene/acetone 5:2) to afford **15-2** (226 mg, 211 μmol, 35%) as a pale, colorless oil.

Synthesis of 3,6,9,12,15,18-Hexaoxatriacontanyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-acetyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-acetyl-α-D-mannopyranoside (15-3)

Compound **15-3** was prepared from **7-3** (292 mg, 163 μmol) according to GP III. The residue was purified by flash column chromatography (silica gel, toluene/acetone 3:1) followed by MPLC (silica gel, toluene/acetone 3:1) to afford **15-3** (132 mg, 97 μmol, 60%) as a pale, colorless oil.

Synthesis of 3,6,9,12,15,18-Hexaoxatriacontanyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-acetyl-α-D-mannopyranosyl-(1→2)-3,4,6-

tri-*O*-acetyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-acetyl-α-D-mannopyranoside (15-4)

Compound **15-4** was prepared from **7-4** (192 mg, 86 μmol) according to GP III. The residue was purified by flash column chromatography (silica gel, toluene/acetone 3:2) followed by MPLC (silica gel, toluene/acetone 3:2) to afford **15-4** (118 mg, 72 μmol, 83%) as a foam.

TLC: R_f = 0.36 (toluene/acetone 3:2); $[\alpha]_D^{20}$ = +29.0 (c = 1.0, $CHCl_3$); ^{1}H-NMR (600 MHz, $CDCl_3$): δ = 0.84 (pseudo t, 3H, C*H*$_3$), 1.17-1.31 (m, 18H, 9C*H*$_2$), 1.54 (pseudo quint, 2H, 1C*H*$_2$), 1.97-2.13 (13s, 39H, 13COC*H*$_3$), 3.41 (pseudo t, 2H, OC*H*$_2$$C_{11}H_{23}$), 3.53-3.55 (m, 2H, OC*H*$_2$), 3.59-3.65 (m, 21H, 21/2OC*H*$_2$), 3.77-3.82 (m, 1H, 1/2OC*H*$_2$), 3.97 (ddd, $^3J_{4a,5a}$ = 9.9 Hz, $^3J_{5a,6a}$ = 4.1 Hz, $^3J_{5a,6'a}$ = 2.5 Hz, 1H, *H*-5a), 4.00-4.23 (m, 14H, *H*-2a, *H*-2b, *H*-2c, *H*-5b, *H*-5c, *H*-5d, *H*-6a, *H*-6b, *H*-6c, *H*-6d, *H*-6´a, *H*-6´b, *H*-6´c, *H*-6´d), 4.94 (d, $^3J_{1d,2d}$ = 1.8 Hz, 1H, *H*-1d), 4.96 (d, $^3J_{1a,2a}$ = 1.9 Hz, 1H, *H*-1a), 5.05 (d, $^3J_{1,2}$ = 2.2 Hz, 1H, *H*-1b or *H*-1c), 5.11 (d, $^3J_{1,2}$ = 2.0 Hz, 1H, *H*-1b or *H*-1c), 5.22-5.35 (m, 4H, *H*-2d, *H*-3a, *H*-3b, *H*-3c, *H*-4a, *H*-4b, *H*-4c, *H*-4d), 5.36 (dd, $^3J_{2d,3d}$ = 3.4 Hz, $^3J_{3d,4d}$ = 10.0 Hz, 1H, *H*-3d) ppm; ^{13}C-NMR (151 MHz, $CDCl_3$): δ = 14.03 (1C, *C*H_3), 20.52-20.78 (13C, 13CO*C*H_3), 22.59-31.81 (10C, 10*C*H_2), 61.97-62.37 (4C, *C*-6a, *C*-6b, *C*-6c, *C*-6d), 65.99-66.38 (4C, *C*-4a, *C*-4b, *C*-4c, *C*-4d), 67.34 (1C, 1O*C*H_2), 68.27 (1C, *C*-3d), 68.39 (1C, *C*-5a), 69.13-69.39 (3C, *C*-5b, *C*-5c, *C*-5d), 69.49 (1C, *C*-2d), 69.59 (1C, *C*-3b or *C*-3c), 69.71 (1C, *C*-3b or *C*-3c), 69.88-69.94 (2C, 2O*C*H_2), 70.33 (1C, *C*-3a), 70.46-70.57 (9C, 9O*C*H_2), 71.45 (1C, O*C*$H_2C_{11}H_{23}$), 76.42-77.13 (3C, *C*-2a, *C*-2b, *C*-3c), 98.31 (1C, *C*-1a), 99.15 (1C, *C*-1d), 99.55 (1C, *C*-1b or *C*-1c), 99.73 (1C, *C*-1b or *C*-1c), 169.20-170.75 (13C, 13*C*OCH_3) ppm; elemental analysis is calcd. (%) for $C_{74}H_{116}O_{40}$ (1645.7): C: 54.01, H: 7.10; found: C: 53.95, H: 7.43; MALDI-MS (pos. mode, CHCA): $[M+Na]^+$ calcd.: 1668.7, found: 1668.7; $[M+K]^+$ calcd.: 1684.8, found: 1684.3.

Synthesis of 3,6,9,12,15,18-Hexaoxatriacontanyl 2,3,4,6-tetra-*O*-acetyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-acetyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-acetyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-acetyl-α-D-mannopyranosyl-(1→2)-3,4,6-tri-*O*-acetyl-α-D-mannopyranoside (15-5)

Compound **15-5** was prepared from **7-5** (225 mg, 85 mmol) according to GP III. The residue was purified by flash column chromatography (silica gel, toluene/acetone 3:2) followed by MPLC (silica gel, toluene/acetone 3:2) to afford **15-5** (116 mg, 60 mmol, 71%) as a foam.

Synthesis of Methyl α-D-mannopyranosyl-(1→2)-α-D-mannopyranoside (16-2)[2]

Compound **16-2** (58 mg, 163 µmol) was prepared from **12-2** (110 mg, 169 µmol) according to GP II as a colorless lyophilizate.

Synthesis of Methyl α-D-mannopyranosyl-(1→2)-α-D-mannopyranosyl-(1→2)-α-D-mannopyranoside (16-3)[3]

Compound **16-3** (43 mg, 83 µmol) was prepared from **12-3** (81 mg, 86 µmol) according to GP II as a colorless lyophilizate.

Triethylene glycol (11.2 g, 75.2 mmol), tritylchloride (20.9, 75.0 mmol) and DMAP (1.0 g, 8.2 mmol) were dissolved in dry pyridine (200 ml) and stirred at room temperature for 16 h. The reaction mixture was concentrated under reduced pressure to 1/3 of the original volume. Precipitated salts were removed by filtration over cellite. The filtrate was reduced and purified by flash column chromatography (silica gel, toluene/acetone 10:1 → 2:1) to afford compound **20** (15.3 g, 39.0 mmol, 52%) as a oil.

Hexaethylene glycol (15.0 g, 53.1 mmol), tritylchloride (14.8 g, 53.1 mmol) and DMAP (650 mg, 5.32 mmol) were dissolved in dry pyridine (400 ml) and stirred at room temperature for 16 h. The reaction mixture was concentrated under reduced pressure to 1/3 of the original volume. Precipitated salts were removed by filtration over cellite. The filtrate was reduced and purified by flash column chromatography (silica gel, toluene/acetone 2:1 → 1:1 → 0:1) to afford compound **21** (13.1 g, 25.0 mmol, 47%) as a oil and not consumed hexaethylene glycol (5.5 g, 19.5 mmol, 37%).

Compound **20** (3.10 g, 7.90 mmol) and TBAI (200 mg, 0.54 mmol) were dissolved in dry THF (80 ml) and NaH (300 mg, 1.25 mol) is added in small portions. After the H_2 development had stopped dodecyl bromide (2.30 ml, 2.39 g, 9.60 mmol) was added. The reaction mixture is refluxed for 16 h. After cooling to RT the precipitated salts were removed by filtration. The filtrate was concentrated under reduced pressure and the residue was purified by flash column chromatography (silica gel, toluene/ethyl acetate 20:1 → 10:1) to afford compound **22** (3.52 g, 6.28 mmol, 79%) as a oil.

TLC: R_f = 0.58 (toluene/ethyl acetate 10:1); ^{1}H-NMR (400 MHz, $CDCl_3$): δ = 0.85-0.92 (m, 3H, CH_3), 1.20-1.39 (m, 18H, 9CH_2), 1.52-1.60 (m, 2H, 1CH_2), 3.20-3.28 (m, 2H, 1OCH_2), 3.40-3.48 (m, 2H, 1OCH_2), 3.52-3.75 (m, 10H, 5OCH_2), 7.18-7.49 (m, 15H, Ph) ppm; $C_{37}H_{52}O_4$ (560.8); MALDI-MS (pos. mode, DHB): $[M+Na]^+$ calcd.: 583.4, found: 583.5; $[M+K]^+$ calcd.: 599.5, found: 599.5.

Compound **21** (13.0 g, 24.8 mmol) and TBAI (900 mg, 2.44 mmol) were dissolved in dry THF (200 ml) and NaH (1.3 g (60%), 32.5 mmol) is added in small portions. After the H_2 development had stopped dodecyl bromide (7.70 ml, 8.01 g, 32.1 mmol) was added. The reaction mixture is refluxed for 16 h. After cooling to RT the precipitated salts were removed by filtration. The filtrate was concentrated under reduced pressure and the residue was purified by flash column chromatography (silica gel, toluene/acetone 2:1) to afford compound **23** (14.4 g, 20.8 mmol, 84%) as a oil .

TLC: R_f = 0.58 (toluene/acetone 5:1) $[\alpha]_D^{20}$ = +6.0 (c = 1.0, CHCl$_3$); ^{1}H-NMR (600 MHz, CDCl$_3$): δ = 0.88 (pseudo t, 3H, CH_3), 1.21-1.34 (m, 18H, 9CH_2), 1.56 (pseudo quint, 2H, CH_2), 2.14 (s, 3H, COCH_3), 3.42 (pseudo t, 2H, OCH_2 C$_{11}$H$_{23}$), 3.54-3.57 (m, 2H, 1OCH_2), 3.60-3.65 (m, 8H, 4OCH_2), 3.70 (dd, $^3J_{5,6}$ = 1.3 Hz, $^2J_{6,6'}$ = 10.3 Hz, 1H, *H*-6), 3.77-3.84 (m, 3H, *H*-5, *H*-6´, 1/2OCH_2), 3.89 (dd, 1H, *H*-4), 3.99 (dd, $^3J_{2,3}$ = 3.3 Hz, $^3J_{3,4}$ = 9.3 Hz, 1H, *H*-3), 4.46-4.60 (m, 6H, 3PhCH_2), 4.88 (d, $^3J_{1,2}$ = 1.8 Hz, 1 H, *H*-1), 5.41 (dd, $^3J_{1,2}$ = 1.8 Hz, $^3J_{2,3}$ = 3.3 Hz, 1H, *H*-2), 7.14-7.36 (m, 9H, *Ph*) ppm; ^{13}C-NMR (151 MHz, CDCl$_3$): δ = 14.11 (1C, CH$_3$), 21.12 (1C, COCH$_3$), 22.67-31.90 (10C, 10CH$_2$), 66.86 (1C, 1OCH$_2$), 68.68 (1C, *C*-2), 68.84 (1C, *C*-6),70.01-70.67 (5C, 5OCH$_2$), 71.35 (1C, *C*-5), 71.53 (OCH$_2$C$_{11}$H$_{23}$), 71.79-73.43 (2C, 2CH$_2$Ph), 74.29 (1C, *C*-4), 75.16 (1C, 1CH$_2$Ph), 78.19 (1C, *C*-3), 97.87 (1C, *C*-1), 127.56-138.38 (18C, *Ph*), 170.41 (1C, COCH$_3$), ppm; $C_{47}H_{68}O_{10}$ (793.0); MALDI-MS (pos. mode, CHCA): [M+Na]$^+$ calcd.: 815.5, found: 815.5; [M+K]$^+$ calcd.: 831.5, found: 831.5.

TLC: R_f = 0.68 (toluene/acetone 5:1) $[\alpha]_D^{20}$ = +14.4 (c = 1.0, CHCl$_3$); ^{1}H-NMR (600 MHz, CDCl$_3$): δ = 0.89 (pseudo t, 3H, CH_3), 1.20-1.33 (m, 18H, 9CH_2), 1.55 (pseudo quint, 2H, 1CH_2), 2.13 (s, 3H, COCH_3), 3.42 (pseudo t, 2H, OCH_2C$_{11}$H$_{23}$), 3.46-3.52 (m, 1H, 1/2OCH_2); 3.53-3.81 (m, 16H, *H*-5a or *H*-5b, *H*-6a, *H*-6b, *H*-6´a, *H*-6´b, 11/2OCH_2), 3.82-3.89 (m, 2H, *H*-4a, *H*-4b), 3.92-4.01 (m, 3H, *H*-3a, *H*-3b, *H*-5a or *H*-5b), 4.06 (dd, 1H, *H*-2a), 4.39-4.86 (m, 12H, 6PhCH_2), 4.91 (d, $^3J_{1a,2a}$ = 1.3 Hz, 1H, *H*-1a), 5.10 (d, $^3J_{1b,2b}$ = 1.8 Hz, 1H, *H*-1b), 5.55 (dd, $^3J_{1b,2b}$ = 1.8 Hz, $^3J_{2b,3b}$ = 2.7 Hz, 1H, *H*-2b), 7.14-7.37 (m, 30H, *Ph*) ppm; ^{13}C-NMR (151 MHz, CDCl$_3$): δ = 14.08 (1C, CH$_3$), 21.09 (1C, COCH$_3$), 22.64-31.87 (10C, 10CH$_2$), 66.62 (1C, 1OCH$_2$), 68.67 (1C, *C*-2b), 68.96 (1C, *C*-6a or *C*-6b), 69.20 (1C, *C*-6a or *C*-6b), 70.00-70.60 (5C, 5OCH$_2$), 71.50 (1C, OCH$_2$C$_{11}$H$_{23}$), 71.77 (2C, *C*-5a, *C*-5b), 71.85-73.36 (4C, 4CH$_2$Ph), 74.28 (1C, *C*-4a or *C*-4b), 74.56 (1C, *C*-4a or *C*-4b), 74.73 (1C, *C*-2a), 75.00-75.09 (2C, 2CH$_2$Ph), 78.09 (1C, *C*-3b), 79.61 (1C, *C*-3a), 98.71 (1C, *C*-1a), 99.52 (1C, *C*-1b), 127.33-138.45 (36C, *Ph*), 170.06 (1C, COCH$_3$) ppm; elemental analysis is calcd. (%) for $C_{74}H_{96}O_{15}$ (1225.6): C: 72.52, H: 7.90; found: C: 72.11, H: 7.88; MALDI-MS (pos. mode, CHCA): [M+Na]$^+$ calcd.: 1247.7, found: 1247.2; [M+K]$^+$ calcd.: 1263.7, found: 1263.2.

TLC: R_f = 0.25 (toluene/acetone 5:1); $[\alpha]_D^{20}$ = +13.1 (c = 1.0, CHCl$_3$); ^{1}H-NMR (600 MHz, CDCl$_3$): δ = 0.88 (pseudo t, 3H, CH_3), 1.21-1.34 (m, 18H, 9CH_2), 1.57 (pseudo quint, 2H, CH_2), 2.14 (s, 3H, COCH_3), 3.44 (pseudo t, 2H, C$_{11}$H$_{23}$OCH_2), 3.55-3.65 (m, 23H, 23/2OCH_2), 3.70 (d, $^3J_{5,6}$ = 3.7 Hz, $^2J_{6,6'}$ = 10.1 Hz, 1H, *H*-6), 3.76-3.84 (m, 3H, *H*-5, *H*-6´, 1/2OCH_2), 3.89 (dd, 1H, *H*-4), 3.99 (dd, $^3J_{2,3}$ = 3.4 Hz, $^3J_{3,4}$ = 9.3 Hz, 1H, *H*-3), 4.46-4.86 (3d, 6H, 3PhCH_2), 4.88 (d, $^3J_{1,2}$ = 1.8 Hz, 1

H, *H*-1), 5.41 (dd, $^{3}J_{1,2}$ = 1.8 Hz, $^{3}J_{2,3}$ = 3.4 Hz, 1H, *H*-2), 7.14-7.36 (m, 9H, *Ph*) ppm; ^{13}C-NMR (151 MHz, CDCl$_3$): δ = 14.11 (1C, *C*H$_3$), 21.11 (1C, CO*C*H$_3$), 22.68-31.90 (10C, 10*C*H$_2$), 66.84 (1C, 1O*C*H$_2$), 68.67 (1C, *C*-2),68.84 (1C, *C*-6) 70.01-70.64 (11C, 11O*C*H$_2$), 71.36 (1C, *C*-5), 71.54(1C, O*C*H$_2$C$_{11}$H$_{23}$) 71.78-73.43 (2C, 2*C*H$_2$Ph), 74.29 (1C, *C*-4), 75.16 (1C, 1*C*H$_2$Ph), 78.19 (1C, *C*-3), 97.87 (1C, *C*-1), 127.55-138.37 (18C, 3CH$_2$*Ph*), 170.40 (1C, *C*OCH$_3$), ppm; elemental analysis is calcd. (%) for $C_{53}H_{80}O_{13}$ (925.2): C: 68.80, H: 8.72; found: C: 68.59, H: 8.76; MALDI-MS (pos. mode, CHCA): $[M+Na]^+$ calcd.: 947.6, found: 947.7; $[M+K]^+$ calcd.: 963.6, found: 963.7.

TLC: R_f = 0.32 (toluene/acetone 5:1); $[\alpha]_D^{20}$ = +13.7 (c = 1.0, CHCl$_3$); ^{1}H-NMR (600 MHz, CDCl$_3$): δ = 0.88 (pseudo t, 3H, C*H*$_3$), 1.20-1.34 (m, 18H, 9C*H*$_2$), 1.57 (pseudo quint, 2H, 1C*H*$_2$), 2.12 (s, 3H, COC*H*$_3$), 3.42-3.50 (m, 3H, 1/2OC*H*$_2$, OC*H*$_2$C$_{11}$H$_{23}$), 3.53-3.81 (m, 28H, *H*-5a or *H*-5b, *H*-6a, *H*-6b, *H*-6´a, *H*-6´b, 23/2OC*H*$_2$), 3.82-3.87 (m, 2H, *H*-4a, *H*-4b), 3.91-3.99 (m, 3H, *H*-3a, *H*-3b, *H*-5a or *H*-5b), 4.05 (dd, 1H, *H*-2a), 4.37-4.86 (m, 12H, 6PhC*H*$_2$), 4.89 (d, $^{3}J_{1a,2a}$ = 1.3 Hz, 1H, *H*-1a), 5.09 (d, 1H, *H*-1b), 5.54 (dd, $^{3}J_{1b,2b}$ = 1.8 Hz, $^{3}J_{2b,3b}$ = 2.7 Hz, 1H, *H*-2b), 7.13-7.36 (m, 30H, *Ph*) ppm; ^{13}C-NMR (151 MHz, CDCl$_3$): δ = 14.10 (1C, *C*H$_3$), 21.12 (1C, CO*C*H$_3$), 22.66-31.88 (10C, 10*C*H$_2$), 66.60 (1C, 1O*C*H$_2$), 68.64 (1C, *C*-2b), 68.92 (1C, *C*-6a or *C*-6b), 69.17(1C, *C*-6a or *C*-6b), 69.00-70.58 (11C, 11O*C*H$_2$), 71.52 (1C, O*C*H$_2$C$_{11}$H$_{23}$), 71.75 (2C, *C*-5, *C*-5b), 71.85-73.36 (4C, 4*C*H$_2$Ph), 74.25 (1C, *C*-4a or *C*-4b), 74.54 (1C, *C*-4a or *C*-4b), 74.68 (1C, *C*-2a), 75.02-75.12 (2C, 2*C*H$_2$Ph), 78.09 (1C, *C*-3b), 79.60 (1C, *C*-3a), 98.71 (1C, *C*-1a), 99.52 (1C, *C*-1b), 127.32-138.42 (36C, *Ph*), 170.08 (1C, *C*OCH$_3$) ppm; $C_{80}H_{108}O_{18}$ (1356.7); MALDI-MS (pos. mode, CHCA): $[M+Na]^+$ calcd.: 1379.8, found: 1379.6; $[M+K]^+$ calcd.: 1395.8, found: 1395.2.

TLC: R_f = 0.41 (toluene/acetone 5:1); $[\alpha]_D^{20}$ = +17.8 (c = 1.0, CHCl$_3$); ^{1}H-NMR (600 MHz, CDCl$_3$): δ = 0.89 (pseudo t, 3H, C*H*$_3$), 1.21-1.35 (m, 18H, 9C*H*$_2$), 1.58 (pseudo quint, 2H, 1C*H*$_2$), 2.14 (s, 3H, COC*H*$_3$), 3.43-3.47 (m, 3H, OC*H*$_2$C$_{11}$H$_{23}$, 1/2OC*H*$_2$), 3.51-3.84 (m, 32H, *H*-4a, *H*-4b, *H*-5a, *H*-6a, *H*-6b, *H*-6c, *H*-6´a, *H*-6´b, *H*-6´c, 23/2OC*H*$_2$), 3.86-3.96 (m, 5H, *H*-3a, *H*-3b, *H*-4c, *H*-5b, *H*-5c), 4.00 (dd, $^{3}J_{2c,3c}$ = 3.3 Hz, $^{3}J_{3c,4c}$ = 9.3 Hz, 1H, *H*-3c), 4.03 (dd, 1H, *H*-2a), 4.10 (dd, 1H, *H*-2b), 4.30-4.86 (m, 18H, 9PhC*H*$_2$), 4.94 (d, $^{3}J_{1a,2a}$ = 1.8 Hz, 1H, *H*-1a), 5.05 (d, $^{3}J_{1c,2c}$ = 1.8 Hz, 1H, *H*-1c), 5.20 (d, $^{3}J_{1b,2b}$ = 1.8 Hz, 1H, *H*-1b), 5.54 (dd, $^{3}J_{1c,2c}$ = 1.8 Hz, $^{3}J_{2c,3c}$ = 3.3 Hz, 1H, *H*-2c), 7.13-7.36 (m, 45H, *Ph*) ppm; ^{13}C-NMR (151 MHz, CDCl$_3$): δ = 14.10 (1C, *C*H$_3$), 21.13 (1C, CO*C*H$_3$), 22.66-31.88 (10C, 10*C*H$_2$), 66.65(1C, 1O*C*H$_2$), 68.69 (1C, *C*-2c), 68.72-69.45 (3C, *C*-6a, *C*-6b, *C*-6c),70.01-70.58 (11C, 11O*C*H$_2$), 71.52 (1C, O*C*H$_2$C$_{11}$H$_{23}$), 71.78-72.16 (6C, *C*-5a, *C*-5b, *C*-5c, 3*C*H$_2$Ph), 73.22-73.32 (3C, 3*C*H$_2$Ph), 74.19 (1C, *C*-4c), 74.67 (1C, *C*-4b), 74.79 (1C, *C*-4a), 74.86-75.08 (6C, *C*-2a, *C*-2b, 4*C*H$_2$Ph), 78.09 (1C, *C*-3b), 79.27 (1C, *C*-3b), 79.44 (1C, *C*-3a), 98.83 (1C, *C*-1a), 99.36 (1C, *C*-1c), 100.57 (1C, *C*-1b), 127.34-138.54 (54C, *Ph*), 170.09 (1C, *C*OCH$_3$) ppm; $C_{107}H_{136}O_{23}$ (1790.2); MALDI-MS (pos. mode, CHCA): $[M+Na]^+$ calcd.: 1813.2, found: 1812.8; $[M+K]^+$ calcd.: 1829.3, found: 1828.6.

TLC: R_f = 0.47 (toluene/acetone 5:1); $[\alpha]_D^{20}$ = +15.6 (c = 1.0, CHCl$_3$); ^{1}H-NMR (600 MHz, CDCl$_3$): δ = 0.89 (pseudo t, 3H, C*H*$_3$), 1.21-1.35 (m, 18H, 9C*H*$_2$), 1.58 (pseudo quint, 2H, 1C*H*$_2$), 2.13 (s, 3H, COC*H*$_3$), 3.40-3.98 (m, 50H, *H*-3a, *H*-3b, *H*-3c, *H*-3d, *H*-4a, *H*-4b, *H*-4c, *H*-4d, *H*-4e, *H*-5a, *H*-5b, *H*-5c, *H*-5d, *H*-5e, *H*-6a, *H*-6b, *H*-6c, *H*-6d, *H*-6e, *H*-6´a, *H*-6´b, *H*-6´c, *H*-6´d, *H*-6´e, 12OC*H*$_2$, OC*H*$_2$C$_{11}$H$_{23}$), 3.99-4.03 (m, 2H, *H*-2a, *H*-3e), 4.07-4.22 (m, 6H, *H*-2b, *H*-2c, *H*-2d, 3/2PhC*H*$_2$), 4.33-4.90 (m, 27H, 27/2PhC*H*$_2$), 4.98 (d, $^3J_{1a,2a}$ = 1.7 Hz, 1H, *H*-1a), 5.04 (d, $^3J_{1e,2e}$ = 1.8 Hz, 1H, *H*-1e), 5.16 (d, $^3J_{1d,2d}$ = 1.7 Hz, 1H, *H*-1d), 5.23 (d, $^3J_{1,2}$ = 1.6 Hz, 1H, *H*-1b or *H*-1c), 5.24 (d, $^3J_{1,2}$ = 1.5 Hz, 1H, *H*-1b or *H*-1c), 5.56 (dd, $^3J_{1e,2e}$ = 1.8 Hz, $^3J_{2e,3e}$ = 3.2 Hz, 1H, *H*-2e), 6.96-7.36 (m, H, 75*Ph*) ppm; ^{13}C-NMR (151 MHz, CDCl$_3$): δ = 14.11 (1C, *C*H$_3$), 21.15 (1C, CO*C*H$_3$), 22.67-31.90 (10C, 10*C*H$_2$), 66.65 (1C, 1O*C*H$_2$), 68.63-69.61 (6C, *C*-2e, *C*-6a, *C*-6b, *C*-6c, *C*-6d, *C*-6e), 70.02-70.60 (11C, 11O*C*H$_2$), 71.53 (1C, O*C*H$_2$C$_{11}$H$_{23}$), 71.63-72.28 (10C, *C*-5a, *C*-5b, *C*-5c, *C*-5d, *C*-5e, 5*C*H$_2$Ph), 73.21-73.32 (5C, 5*C*H$_2$Ph), 74.24 (1C, *C*-4e), 74.70-75.15 (10C, *C*-2d, *C*-4a, *C*-4b, *C*-4c, *C*-4d, 5*C*H$_2$Ph), 75.56 (1C, *C*-2b or *C*-2c), 75.79 (1C, *C*-2a), 76.18 (1C, *C*-2b or *C*-2c), 78.18 (1C, *C*-3e), 79.00-79.26 (3C, *C*-3a, *C*-3b, *C*-3c), 79.42 (1C, 1*C*-d), 98.84 (1C, *C*-1a), 99.38 (1C, *C*-1e), 100.73 (1C, *C*-1d), 101.30 (1C, *C*-1b or *C*-1c), 101.33 (1C, *C*-1b or *C*-1c), 127.27-138.63 (90C, *Ph*), 170.05 (1C, *C*OCH$_3$) ppm; elemental analysis is calcd. (%) for $C_{161}H_{192}O_{33}$ (2655.2): C: 72.83, H: 7.29; found: C: 72.30, H: 7.53; MALDI-MS (pos. mode, CHCA): [M+Na]$^+$ calcd.: 2678.2, found: 2677.4; [M+K]$^+$ calcd.: 2694.3, found: 2694.4.

TLC: R$_f$ = 0.41 (toluene/acetone 2:1); $[\alpha]_D^{20}$ = +23.3 (c = 1.0, CHCl$_3$); ^{1}H-NMR (600 MHz, CDCl$_3$): δ = 1.98-2.13 (10s, 30 H, 10COC*H*$_3$), 3.39 (s, 3H, OC*H*$_3$), 3.89 (ddd, $^3J_{4a,5a}$ = 9.6 Hz, $^3J_{5a,6a}$ = 4.5 Hz, $^3J_{5a,6'a}$ = 2.4 Hz, 1H, *H*-5a), 4.01 (dd, 1 H, *H*-2a), 4.05-4.24 (m, 9H, *H*-2b, *H*-5b, *H*-5c, *H*-6a, *H*-6´a, *H*-6b, *H*-6´b, *H*-6c, *H*-6´c), 4.84 (d, $^3J_{1a,2a}$ = 2.0 Hz, 1 H, *H*-1a), 4.92 (d, $^3J_{1c,2c}$ = 1.8 Hz, 1 H, *H*-1c), 5.08 (d, $^3J_{1b,2b}$ = 2.0 Hz, 1H, *H*-1b), 5.22-5.33 (m, 6H, *H*-2c, *H*-3a, *H*-3b, *H*-4a, *H*-4b, *H*-4c), 5.37 (dd, $^3J_{2c,3c}$ = 3.4 Hz, $^3J_{3c,4c}$ = 10.0 Hz, 1H, *H*-3c) ppm; ^{13}C-NMR (151 MHz, CDCl$_3$): δ = 20.57-20.78 (10C, 10CO*C*H$_3$), 55.19 (1C, O*C*H$_3$), 62.09-62.54 (3C, *C*-6a, *C*-6b, *C*-6c), 66.16-66.26 (3C, *C*-4a, *C*-4b, *C*-4c), 68.34 (1C, *C*-3c), 68.44 (1C, *C*-5a), 69.21 (1C, *C*-5b or *C*-5c), 69.42 (1C, *C*-5b or *C*-5c), 69.51-70.37 (3C, *C*-2c, *C*-3a, *C*-3b), 76.56 (1C, *C*-2a), 77.36 (1C, *C*-2b), 99.18 (1C, *C*-1a), 99.33 (1C, *C*-1c), 99.75 (1C, *C*-1b), 169.33-170.81 (10C, 10*C*OCH$_3$) ppm; $C_{39}H_{54}O_{26}$ (938.8); MALDI-MS (pos. mode, CHCA): [M+Na]$^+$ calcd.: 961.3, found: 962.1; [M+K]$^+$ calcd.: 977.3, found: 978.1.

TLC: R$_f$ = 0.54 (toluene/acetone 3:1); $[\alpha]_D^{20}$ = +24.0 (c = 1.0, CHCl$_3$); ^{1}H-NMR (600 MHz, CDCl$_3$): δ = 0.87 (pseudo t, 3H, C*H*$_3$), 1.19-1.37 (m, 18H, 9C*H*$_2$), 1.58 (pseudo quint, 2H, 1C*H*$_2$), 2.00-2.14 (7s, 21H, 7COC*H*$_3$), 3.39-3.44 (m, 1H, 1/2OC*H*$_2$C$_{11}$H$_{23}$), 3.64-3.69 (m, 1H, 1/2OC*H*$_2$C$_{11}$H$_{23}$), 3.90 (ddd, $^3J_{4a,5a}$ = 9.3 Hz, $^3J_{5a,6a}$ = 4.2 Hz, $^3J_{5a,6'a}$ = 2.3 Hz, 1H, *H*-5a), 4.01 (dd, 1H, *H*-2a), 4.08-4.24 (m, 5H, *H*-5b, *H*-6a, *H*-6´a, *H*-6b, *H*-6´b), 4.90-4.92 (2d, 2H, *H*-1a, *H*-1b), 5.24-5.34 (m, 4H, *H*-2b, *H*-3a, *H*-4a, *H*-4b), 5.40 (dd, $^3J_{2b,3b}$ = 3.4 Hz, $^3J_{3b,4b}$ = 10.0 Hz, 1H, *H*-3b) ppm; ^{13}C-NMR (151 MHz, CDCl$_3$): δ = 14.10 (1C, *C*H$_3$), 20.66-20.87 (7C, 7CO*C*H$_3$), 22.66-31.88 (10C, 10*C*H$_2$), 62.19 (1C, *C*-6a), 62.52 (1C, *C*-6b), 66.23 (1C, *C*-4a), 66.34 (1C, *C*-4b), 68.35 (1C, *C*-5a or *C*-3b), 68.42 (1C, *C*-5a or *C*-3b), 68.53 (1C, O*C*H$_2$C$_{11}$H$_{23}$), 69.07 (1C, *C*-5b), 69.75 (1C, *C*-2b), 70.34 (1C, *C*-3a), 77.18 (1C, *C*-2a), 98.22 (1C, *C*-1a or *C*-1b), 99.13 (1C, *C*-1a or *C*-1b), 169.42-170.88 (7C, 7*C*OCH$_3$) ppm; $C_{38}H_{60}O_{18}$ (804.9); MALDI-MS (pos. mode, CHCA): [M+Na]$^+$ calcd.: 827.4, found: 827.3; [M+K]$^+$ calcd.: 843.4, found: 843.3.

TLC: R_f = 0.55 (toluene/acetone 3:2); $[\alpha]_D^{20}$ = +21.7 (c = 1.0, $\mathrm{CHCl_3}$); ^{1}H-NMR (600 MHz, $\mathrm{CDCl_3}$): δ = 0.87 (pseudo t, 3H, $\mathrm{C}H_3$), 1.20-1.33 (m, 18H, 9$\mathrm{C}H_2$), 1.56 (pseudo quint, 2H, 1$\mathrm{C}H_2$), 1.97-2.14 (4s, 12H, 4$\mathrm{COC}H_3$), 3.44 (pseudo t, 2H, $\mathrm{O}\mathrm{C}H_2\mathrm{C_{11}H_{23}}$), 3.55-3.57 (m, 2H, 1$\mathrm{OC}H_2$), 3.61-3.68 (m, 21H, 21/2$\mathrm{OC}H_2$), 3.77-3.82 (m, 1H, 1/2$\mathrm{OC}H_2$), 4.03-4.10 (m, 2H, *H*-5, *H*-6), 4.28 (dd, $^3J_{5,6'}$ = 4.9 Hz, $^2J_{6,6'}$ = 12.2 Hz, 1H, *H*-6´), 4.86 (d, $^3J_{1,2}$ = 1.7 Hz, 1H, *H*-1), 5.25 (dd, $^3J_{1,2}$ = 1.7 Hz, $^3J_{2,3}$ = 3.5 Hz, 1H, *H*-2), 5.27 (dd, 1H, *H*-4), 5.34 (dd, $^3J_{2,3}$ = 3.5 Hz, $^3J_{3,4}$ = 10.0 Hz, 1H, *H*-3) ppm; ^{13}C-NMR (151 MHz, $\mathrm{CDCl_3}$): δ = 14.08 (1C, $C\mathrm{H_3}$), 20.65-20.86 (4C, 4$\mathrm{CO}C\mathrm{H_3}$), 22.64-31.87 (10C, 10$C\mathrm{H_2}$), 62.37 (1C, *C*-6), 66.11 (1C, *C*-4), 67.34 (1C, 1$\mathrm{O}C\mathrm{H_2}$), 68.35 (1C, *C*-5), 69.05 (1C, *C*-3), 69.53 (1C, *C*-2), 69.94-70.67 (11C, 11$\mathrm{O}C\mathrm{H_2}$), 71.51 (1C, $\mathrm{O}C\mathrm{H_2C_{11}H_{23}}$), 97.68 (1C, *C*-1), 169.68-170.62 (4C, 4$C\mathrm{OCH_3}$) ppm; elemental analysis is calcd. (%) for $\mathrm{C_{38}H_{68}O_{16}}$ (780.9): C: 58.44, H: 8.78; found: C: 58.25, H: 9.34; MALDI-MS MALDI-MS (pos. mode, CHCA): $[M+Na]^+$ calcd.: 803.5, found: 803.3; $[M+K]^+$ calcd.: 819.5, found: 819.2.

TLC: R_f = 0.51 (toluene/acetone 3:2); $[\alpha]_D^{20}$ = +19.9 (c = 1.0, $\mathrm{CHCl_3}$); ^{1}H-NMR (600 MHz, $\mathrm{CDCl_3}$): δ = 0.83 (pseudo t, 3H, $\mathrm{C}H_3$), 1.16-1.30 (m, 18H, 9$\mathrm{C}H_2$), 1.53 (pseudo quint, 2H, 1$\mathrm{C}H_2$), 1.97-2.11 (7s, 21H, 7$\mathrm{COC}H_3$), 3.40 (pseudo t, 2H, $\mathrm{OC}H_2\mathrm{C_{11}H_{23}}$), 3.52-3.54 (m, 2H, $\mathrm{OC}H_2$), 3.58-3.64 (m, 21H, 21/2$\mathrm{OC}H_2$), 3.75-3.81 (m, 1H, 1/2$\mathrm{OC}H_2$), 3.95 (ddd, 1H, *H*-5a), 4.02 (dd, 1H, *H*-2a), 4.05-4.14 (m, 3H, *H*-5b, *H*-6a, *H*-6b), 4.17-4.21 (m, 2H, *H*-6´a, *H*-6´b), 4.88 (d, 1H, *H*-1b), 4.94 (d, 1H, *H*-1a), 5.21-5.33 (m, 4H, *H*-2b, *H*-3a, *H*-4a, *H*-4b), 5.37 (dd, $^3J_{2b,3b}$ = 3.3 Hz, $^3J_{3b,4b}$ = 10.0 Hz, 1H, *H*-3b) ppm; ^{13}C-NMR (151 MHz, $\mathrm{CDCl_3}$): δ = 14.01 (1C, $C\mathrm{H_3}$), 20.56-20.77 (7C, 7$\mathrm{CO}C\mathrm{H_3}$), 22.56-31.78 (10C, 10$C\mathrm{H_2}$), 61.95 (1C, *C*-6a or *C*-6b), 62.39 (1C, *C*-6a or *C*-6b), 66.01 (1C, *C*-4a), 66.25 (1C, *C*-4b), 67.29(1C, 1$\mathrm{O}C\mathrm{H_2}$), 68.26 (1C, *C*-3b), 68.33 (1C, *C*-5a), 69.00 (1C, *C*-5b), 69.64 (1C, *C*-2b), 69.85-69.91 (2C, 2$\mathrm{O}C\mathrm{H_2}$), 70.15 (1C, *C*-3a), 70.44-70.56 (9C, 9$\mathrm{O}C\mathrm{H_2}$), 71.41 (1C, $\mathrm{O}C\mathrm{H_2C_{11}H_{23}}$), 76.97 (1C, *C*-2a), 98.23 (1C, *C*-1a), 99.04 (1C, *C*-1b), 169.27-170.79 (7C, 7$C\mathrm{OCH_3}$) ppm; $\mathrm{C_{50}H_{84}O_{24}}$ (1069.2); MALDI-MS (pos. mode, CHCA): $[M+Na]^+$ calcd.: 1091.5, found: 1090.9; $[M+K]^+$ calcd.: 1107.5, found: 1106.9.

TLC: R_f = 0.43 (toluene/acetone 3:2); $[\alpha]_D^{20}$ = +15.9 (c = 1.0, $\mathrm{CHCl_3}$); ^{1}H-NMR (600 MHz, $\mathrm{CDCl_3}$): δ = 0.85 (pseudo t, 3H, $\mathrm{C}H_3$), 1.17-1.31 (m, 18H, 9$\mathrm{C}H_2$), 1.54 (pseudo quint, 2H, 1$\mathrm{C}H_2$), 1.97-2.12 (10s, 30H, 10$\mathrm{COC}H_3$), 3.41 (pseudo t, 2H, $\mathrm{OC}H_2\mathrm{C_{11}H_{23}}$), 3.53-3.56 (m, 2H, 1$\mathrm{OC}H_2$), 3.59-3.65 (m, 21H, 21/2$\mathrm{OC}H_2$), 3.77-3.83 (m, 1H, 1/2$\mathrm{OC}H_2$), 3.97 (ddd, $^3J_{4a,5a}$ = 9.1 Hz, $^3J_{5a,6a}$ = 4.0 Hz, $^3J_{5a,6'a}$ = 2.4 Hz, 1H, *H*-5a), 4.02 (dd, 1H, *H*-2a), 4.05-4.23 (m, 9H, *H*-2b, *H*-5b, *H*-5c, *H*-6a, *H*-6b, *H*-6c, *H*-6´a, *H*-6´b, *H*-6´c), 4.92 (d, $^3J_{1c,2c}$ = 1.8 Hz, 1H, *H*-1c), 4.96 (d, $^3J_{1a,2a}$ = 1.8 Hz, 1H, *H*-1a), 5.08 (d, $^3J_{1b,2b}$ = 2.0 Hz, 1H, *H*-1b), 5.22-5.33 (m, 6H, *H*-2c, *H*-3a, *H*-3b, *H*-4a, *H*-4b, *H*-4c), 5.36 (dd, $^3J_{2c,3c}$ = 3.4 Hz, $^3J_{3c,4c}$ = 10.0 Hz, 1H, *H*-3c) ppm; ^{13}C-NMR (151 MHz, $\mathrm{CDCl_3}$): δ = 14.03 (1C, $C\mathrm{H_3}$), 20.56-20.77 (10C, 10$\mathrm{CO}C\mathrm{H_3}$), 22.59-31.82 (10C, 10$C\mathrm{H_2}$), 61.95-62.43(3C, *C*-6a, *C*-6b, *C*-6c), 66.08-66.20 (3C, *C*-4a, *C*-4b, *C*-4c), 67.33 (1C, 1$\mathrm{O}C\mathrm{H_2}$), 68.31 (1C, *C*-3c), 68.44 (1C, *C*-5a), 69.16 (1C, *C*-5c), 69.37 (1C, *C*-5b), 69.50 (1C, *C*-2c), 69.61 (1C, *C*-3a or *C*-3b), 69.88-69.95 (2C, 2$\mathrm{O}C\mathrm{H_2}$), 70.41 (1C, *C*-3a or *C*-3b), 70.47-70.57 (9C, 9$\mathrm{O}C\mathrm{H_2}$), 71.46 (1C, $\mathrm{O}C\mathrm{H_2C_{11}H_{23}}$), 76.61 (1C, *C*-2a), 77.35 (1C, *C*-2b), 98.27 (1C, *C*-1a), 99.30 (1C, *C*-1c), 99.70 (1C, *C*-1b), 169.27-170.80 (10C, 10$C\mathrm{OCH_3}$) ppm; elemental analysis is calcd. (%) for $\mathrm{C_{62}H_{100}O_{32}}$ (1357.4): C: 54.86, H: 7.43;

found: C: 55.04, H: 7.66; MALDI-MS (pos. mode, CHCA): $[M+Na]^+$ calcd.: 1380.4, found: 1380.1; $[M+K]^+$ calcd.: 1396.5, found: 1396.1.

TLC: R_f = 0.32 (toluene/acetone 3:2); $[\alpha]_D^{20}$ = +28.9 (c = 1.0, $CHCl_3$); ^{1}H-NMR (600 MHz, $CDCl_3$): δ = 0.84 (pseudo t, 3H, CH_3), 1.16-1.31 (m, 18H, 9CH_2), 1.54 (pseudo quint, 2H, 1CH_2), 1.97-2.13 (16s, 48H, 16COCH_3), 3.41 (pseudo t, 2H, OC$H_2C_{11}H_{23}$), 3.53-3.55 (m, 2H, OCH_2), 3.59-3.65 (m, 21H, 21/2OCH_2), 3.77-3.82 (m, 1H, 1/2OCH_2), 3.97 (ddd, $^3J_{4a,5a}$ = 9.8 Hz, $^3J_{5a,6a}$ = 3.9 Hz, $^3J_{5a,6'a}$ = 2.5 Hz, 1H, *H*-5a), 4.01-4.24 (18H, *H*-2a, *H*-2b, *H*-2c, *H*-2d, *H*-5b, *H*-5c, *H*-5d, *H*-5e, *H*-6a, *H*-6b, *H*-6c, *H*-6d, *H*-6e, *H*-6´a, *H*-6´b, *H*-6´c, *H*-6´d, *H*-6´e), 4.94 (d, 1H, *H*-1e), 4.96 (d, $^3J_{1a,2a}$ = 1.6 Hz, 1H, *H*-1a), 5.07 (d, $^3J_{1,2}$ = 2.0 Hz, 1H, *H*-1b or *H*-1c or *H*-1d), 5.09 (d, $^3J_{1,2}$ = 1.9 Hz, 1H, *H*-1b or *H*-1c or *H*-1d), 5.14 (d, $^3J_{1,2}$ = 1.4 Hz, 1H, *H*-1b or *H*-1c or *H*-1d), 5.19-5.35 (m, 9H, *H*-2e, *H*-3a, *H*-3b, *H*-3c, *H*-3d, *H*-4a, *H*-4b, *H*-4c, *H*-4d, *H*-4e), 5.37 (dd, $^3J_{2e,3e}$ = 3.2 Hz, $^3J_{3e,4e}$ = 10.0 Hz, 1H, *H*-3e) ppm; ^{13}C-NMR (151 MHz, $CDCl_3$): δ = 14.02 (1C, *C*H_3), 20.51-20.75 (16C, 16CO*C*H_3), 22.58-31.81 (10C, 10*C*H_2), 61.93-62.34 (5C, *C*-6a, *C*-6b, *C*-6c, *C*-6d, *C*-6e), 65.92-66.37 (5C, *C*-4, *C*-4b, *C*-4c, *C*-4d, *C*-4e), 67.33 (1C, 1O*C*H_2), 68.26 (1C, *C*-3e), 68.40 (1C, *C*-5a), 69.09-69.39 (4C, *C*-5b, *C*-5c, *C*-5d, *C*-5e), 69.50 (1C, *C*-2e), 69.57-69.73 (3C, *C*-3b, *C*-3c, *C*-3d), 69.64 (1C, *C*-2b), 69.87-69.94 (2C, 2O*C*H_2), 70.33 (1C, *C*-3a), 70.46-70.56 (9C, 9O*C*H_2), 71.45 (1C, O*C*$H_2C_{11}H_{23}$), 75.89-77.1 (4C, *C*-2a, *C*-2b, *C*-2c, *C*-2d), 98.33 (1C, *C*-1a), 99.22 (1C, *C*-1e), 99.36 (1C, *C*-1b or *C*-1c or *C*-1d), 99.48 (1C, *C*-1b or *C*-1c or *C*-1d), 99.78 (1C, *C*-1b or *C*-1c or *C*-1d), 169.14-170.75 (16C, 16*C*OCH_3) ppm; elemental analysis is calcd. (%) for $C_{86}H_{132}O_{48}$ (1933.9): C: 53.41, H: 6.88; found: C: 53.64, H: 7.37; MALDI-MS (pos. mode, CHCA): $[M+Na]^+$ calcd.: 1956.3, found: 1956.1; $[M+K]^+$ calcd.: 1972.4, found: 1972.3.

Synthesis of 8-Triphenylmethyloxy-3,6-dioxaoctanol (20)[8]

Synthesis of 17-Triphenylmethoxy-3,6,9,12,15,-pentaoxaheptadecanol (21)[9]

1-Triphenylmethoxy-3,6,9-trioxahenicosan (22)

1-Triphenylmethoxy-3,6,9,12,15,18-hexaoxatriacontan (23)[10]

Literature references for synthetic part:

1. Wrigley, A. N.; Stirton, A. J.; Howard, E.(1960) Higher Alkyl Monoethers of Mono- to Tetraethylene Glycol. *J. Org. Chem.* 25, 439-444.
2. Sallay, P; Morgos, J.; Farkas, L.; Rusznak, I.; Bartha, B (1979) Preparation of Homogeneous Tensides. *Acta Chim Acad Sci Hungar.* 102, (1), 85-90.
3. Beignet, J.; Tiernan, J.; Woo, C. H.; Kariuki, B.; Cox, L. R. (2004) Stereoselective Synthesis of Allyl-*C*-mannosyl Compounds: Use of a Temporary Silicon Connection in Intramolecular Allylation Strategies with Allylsilanes. *J. Org. Chem.* 69, (19), 6341-6356.
4. Peters, T. (1991) Synthesis and Conformational Analysis of Methyl 2-*O*-(α-D-Mannopyranosyl)-α-D-mannopyranoside. *Liebigs Ann. Chem,* (2), 135-141.
5. Ogawa, S.; Furuya, T.; Tsunoda, H., Hindsgaul, O.; Stangier, K., Palcic, M. M. (1995) Synthesis of β-D-Glc*p*NAc-(1→2)-5a-carba-α-D-Man*p*-(1→6)-β-D-Glc*p*-O$(CH_2)_7CH_3$: a reactive acceptor analog for *N*-acetylglucosaminyltransferase-V. *Carbohydr. Res.* 271, 197-205.

6. Rademann, J.; Schmidt, R. R. (1997) Repetitive Solid Phase Glycosylation on an Alkyl Thiol Polymer Leading to Sugar Oligomers Containing 1-2-trans- and 1-2-cis-Glycosidic Linkages. *J. Org. Chem,* 62, 3650-3653.

7. Szurmai Z.; Balatoni L.; Liptak, (1994) A., Synthesis of some partially substituted methyl α-D-and phenyl 1-thio-α-D-mannopyranosides for the preparation of manno-oligosaccharides. *Carbohydr. Res.* 254, 301-309.

8. Kaats-Richters, V. E. M.; Zwikker, J. W.; Keegstra, E. M. D.; Jenneskens, L. W. (1994) Large Scale Monotritylation of Water Soluble Compounds Containing Multiple Hydroxyl Groups. *Synth. Commm.* 24, (17), 2399-2409.

9. Loiseau, F. A.; Hii, K. K. M.; Hill, A. M. (2004) Multigram Synthesis of Well-Defined Extended Bifunctional Polyethylene Glycol (PEG) Chains. *J. Org. Chem,* 69, 639-647.

10. Gingras B. A., Bayley C.H. (1957) Preparation of Some Pure Polyoxyethyleneglycol Ethers. Part I. *Can. J. Chem.* 35, 599-604.

Figure S3: [^{15}N,^{1}H]-HSQC spectra of CVN/DPC in the absence of glycolipids (top), in the presence of **19-2** (middle) and **19-3** (bottom).

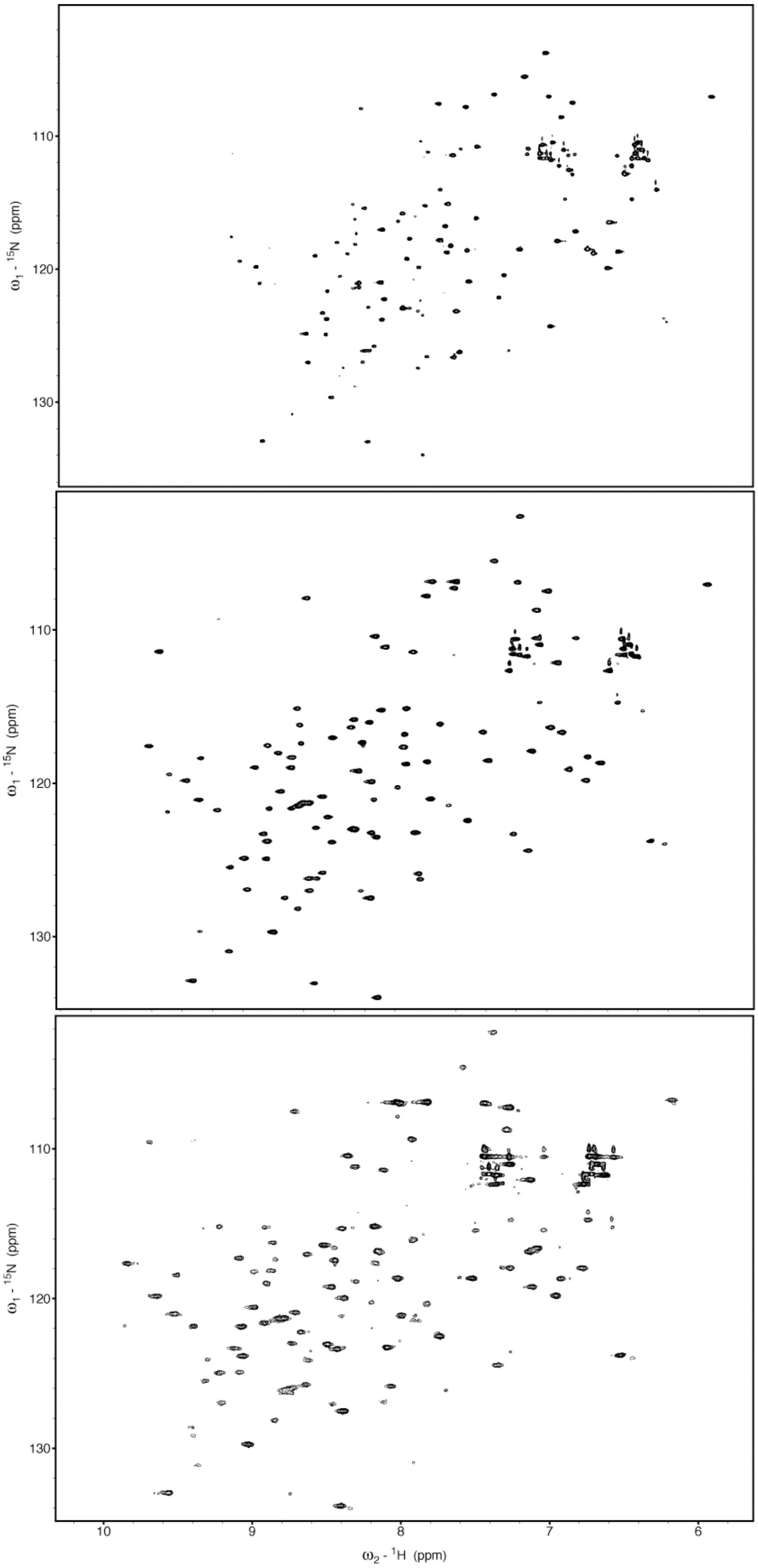


Figure S4: Overlay of [^{15}N,^{1}H]-HSQC spectra of CVN/DPC in the absence of glycolipids (blue contours), in the presence of **19-2** (magenta) and **19-3** (purple).

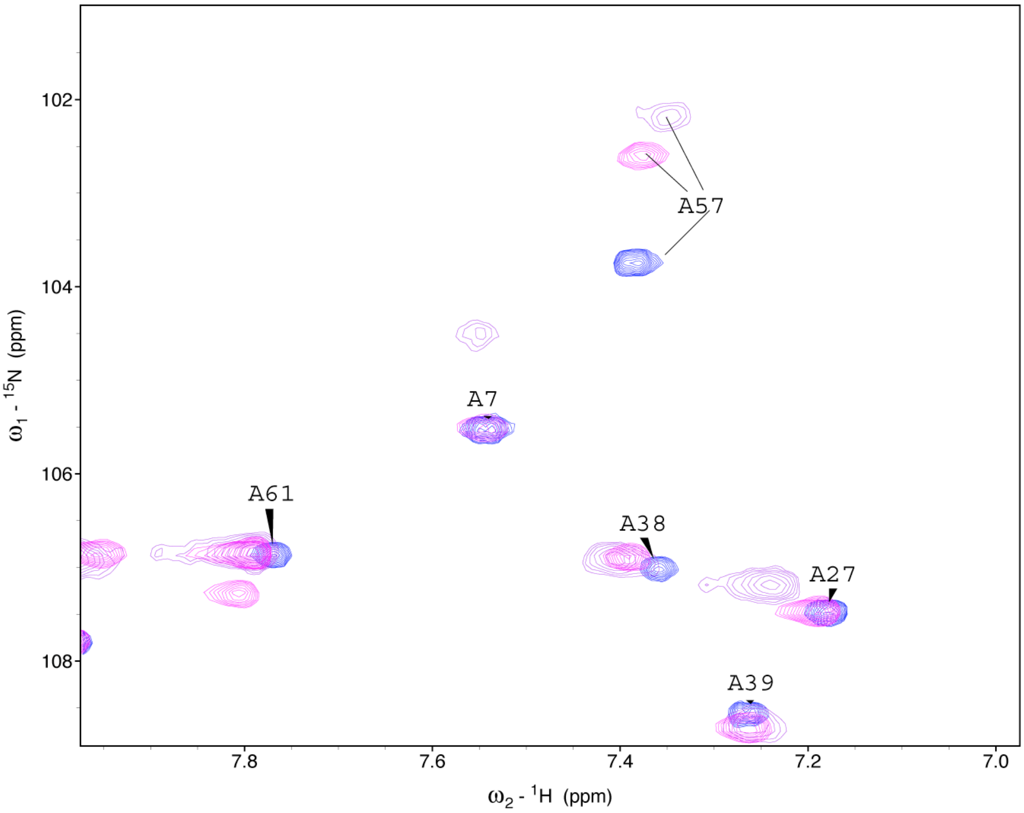


Table S1) Proton and nitrogen chemical shifts of CVN and the triple mutant. Spectra were recorded at approx. 0.3 mM concentration of protein in 300mM DPC / 90% H_2O/10% ^2H_2O at 310 K and pH 6.5 in 50mM sodium phosphate buffer. Chemical shifts are referenced via the proton chemical shift of the water resonance set to 4.63 at 310K. The ^{15}N chemical shift scale was calculated by multiplying the proton frequency of 0 ppm by 0.1013291444 to yield the ^{15}N frequency for 0 ppm.

	Residue		^{15}N	$^{1}H^{N}$	Mutant	
	2	GLY	111.226	9.780		
	3	LYS(N)	119.342	9.716	120.941	8.573
	4	PHE	121.607	9.056	120.571	9.055
	5	SER	116.119	7.939	116.740	8.106
	6	GLN	117.838	7.325	117.937	7.229
	7	THR	105.475	7.567	103.989	7.535
	8	CYS	118.630	6.862	119.169	6.917
	9	TYR	119.774	9.597	119.774	9.604
	10	ASN	115.073	8.871	114.744	8.848
	11	SER	115.047	8.146	115.057	8.170
	12	ALA	124.808	9.220	124.816	9.230
	13	ILE	119.805	8.361	119.753	8.375
	14	GLN	127.377	8.377	127.375	8.368
	15	GLY	118.359	9.494	118.252	9.495
	16	SER	120.494	8.968	120.478	8.966
	17	VAL	123.106	8.084	123.129	8.091
	18	LEU	132.936	8.755	132.897	8.749
	19	THR	123.696	9.068	123.690	9.067
	20	SER	116.985	8.641	117.038	8.669
	21	THR	123.766	8.649	123.387	8.581
	22	CYS	123.250	9.101	123.100	9.030
	23	GLU(I)	122.892	8.493	121.919	8.161
	24	ARG	123.404	8.342	125.122	8.402
	25	THR	118.943	9.147	118.109	8.937
	26	ASN	116.005	8.398	115.940	8.203
	27	GLY	107.434	7.211	107.762	7.562
	28	GLY	107.756	8.013	107.633	7.875
	29	TYR	115.182	8.324	115.575	8.151
	30	ASN	117.924	8.990	117.865	8.957
	31	THR	125.767	8.705	125.897	8.735
	32	SER	122.830	8.751	122.460	8.770
	33	SER	111.381	8.114	111.132	8.141
	34	ILE	121.708	9.359	121.368	9.354
	35	ASP	126.096	8.776	126.385	8.779
	36	LEU	126.970	9.201	126.865	9.205
	37	ASN	117.270	8.841	117.353	8.854
	38	SER	106.968	7.389	107.034	7.411
	39	VAL	108.520	7.289	108.465	7.286
	40	ILE	119.873	6.942	119.863	7.005
	41	GLU	126.522	8.313	125.315	8.324
	42	ASN	120.740	8.410	120.447	8.439
	43	VAL	128.772	8.854	128.557	8.891
	44	ASP	127.983	8.966	127.847	8.982
	45	GLY	107.508	8.217	107.605	8.264
	46	SER	112.832	7.210	112.383	7.252
	47	LEU	124.252	7.378	124.230	7.407
	48	LYS	120.943	8.659	120.765	8.628
	49	TRP	122.209	8.624	122.494	8.647
	50	GLN	126.586	8.101	126.840	8.097
	51					
	Residue				Mutant	
	52	SER	112.190	6.763	112.970	6.855
	53	ASN	110.714	9.408	110.990	9.483
	54	PHE	118.194	8.126	118.194	8.128
	55	ILE	118.764	7.053	118.743	7.052
	56	GLU	118.450	7.096	118.317	7.092
	57	THR	103.698	7.411	103.587	7.412
	58	CYS	117.121	7.182	117.153	7.198
	59	ARG	117.514	9.783	117.493	9.787
	60	ASN	116.193	8.848	116.192	8.852
	61	THR	106.818	7.799	106.877	7.807
	62	GLN	118.790	8.909	118.881	8.913
	63	LEU	122.884	8.433	120.787	7.963
	64	ALA	132.872	9.542	132.685	9.542
	65	GLY	110.330	8.356	110.345	8.357
	66	SER	116.359	8.524	116.204	8.513
	67	SER	111.138	8.301	110.967	8.274
	68	GLU	120.881	7.990	121.925	8.010
69	69	LEU	126.079	8.739	124.389	8.714
70	70	ALA	129.616	9.030	129.645	9.041
	71	ALA	119.170	8.458	119.047	8.462
	72	GLU	117.778	8.205	117.660	8.216
	73	CYS	121.068	9.455	121.005	9.471
	74	LYS	122.309	8.358	122.182	8.357
	75	THR	115.369	8.779	115.254	8.776
	76	ARG	121.026	9.568	120.951	9.534
	77	ALA	120.413	7.720	120.383	7.721
	78	GLN	110.738	7.930	110.749	7.939
	79	GLN	116.427	6.929	116.398	6.935
	80	PHE	121.069	8.875	120.894	8.815
	81	VAL	118.079	8.853	118.079	8.848
	82	SER	117.668	8.439	117.633	8.447
	83	THR	121.381	8.860	121.393	8.825
	84	LYS	122.092	7.767	122.055	7.767
	85	ILE	121.300	8.822	121.115	8.839
	86	ASN	124.870	9.069	124.928	9.056
	87	LEU	125.457	9.317	125.420	9.318
	88	ASP	116.727	8.164	116.425	8.210
	89	ASP	118.544	8.006	117.964	7.978
	90	HIS	106.981	6.157	107.103	6.133
	91	ILE	118.698	8.157	118.850	8.210
	92	ALA	130.864	9.325	131.242	9.136
	93	ASN(A)	115.759	8.486		
	94	ILE	126.950	8.796		
	95	ASP	127.364	8.941	130.009	9.416
	96	GLY	107.886	8.805	102.867	8.582
	97	THR	118.465	7.614	117.858	7.584
	98	LEU	126.198	8.059	126.747	8.263
	99	LYS	120.998	8.820	121.358	9.006
	100	TYR	123.125	8.377	123.252	8.390
	101	GLU	133.930	8.342	133.702	8.364

Table S2: Chemical shift changes of CV-N upon adding the lapidated mannobiose derivatives **17-2**, **18-2**, **19-2** to a solution of 0.2mM CV-N in 300mM DPC:

residue		**17-2**	**18-2**	**19-2**	residue		**17-2**	**18-2**	**19-2**
2	GLY	.060	.067	.075	52	SER	.260	.265	.263
3	LYS	2.152	2.158	2.147	53	ASN	.019	.554	.032
4	PHE	.006	.007	.005	54	PHE	.875	.867	.903
5	SER	.007	.003	.012	55	ILE	.162	.155	.137
6	GLN	.007	.005	.000	56	GLU	.153	.147	.173
7	THR	.007	.023	.004	57	THR	.634	.594	.512
8	CYS	.014	.017	.010	58	CYS	.168	.181	.233
9	TYR	.005	.005	.003	59	ARG	.004	.022	.004
10	ASN	.006	.015	.016	60	ASN	.010	.026	.020
11	SER	.024	.025	.021	61	THR	.013	.021	.016
12	ALA	.009	.016	.005	62	GLN	.025	.061	.045
13	ILE	.017	.021	.022	63	LEU	.074	.065	.070
14	GLN	.020	.024	.021	64	ALA	.014	.022	.020
15	GLY	.019	.026	.021	65	GLY	.017	.026	.023
16	SER	.008	.014	.011	66	SER	.014	.025	.019
17	VAL	.022	.033	.026	67	SER	.014	.030	.020
18	LEU	.014	.026	.021	68	GLU	.011	.038	.027
19	THR	.010	.008	.006	69	LEU	.046	.065	.064
20	SER	.009	.008	.006	70	ALA	.010	.014	.011
21	THR	.013	.014	.012	71	ALA	.010	.022	.009
22	CYS	.004	.002	.011	72	GLU	.070	.084	.087
23	GLU				73	CYS	.073	.065	.065
24	ARG				74	LYS	.499	.513	.462
25	THR				75	THR	.898	.931	.987
26	ASN	.014	.018	.015	76	ARG	.365	.404	.393
27	GLY	.009	.013	.004	77	ALA	.490	.957	1.707
28	GLY	.013	.020	.012	78	GLN	.781	1.214	1.579
29	TYR	.009	.004	.005	79	GLN	.184	.219	.281
30	ASN	.024	.030	.023	80	PHE			
31	THR	.010	.017	.017	81	VAL	.059	.082	.110
32	SER	.006	.010	.006	82	SER	.093	.120	.163
33	SER	.001	.004	.002	83	THR	.070	.076	.101
34	ILE	.028	.029	.020	84	LYS	.129	.133	.127
35	ASP	.030	.038	.037	85	ILE	.076	.041	.074
36	LEU	.012	.045	.035	86	ASN	.003	.007	.003
37	ASN	.015	.034	.035	87	LEU	.012	.015	.010
38	SER	.057	.067	.054	88	ASP	.018	.023	.015
39	VAL	.022	.073	.050	89	ASP	.008	.012	.008
40	ILE	.026	.062	.047	90	HIS	.025	.020	.012
41	GLU	.245	.241	.235	91	ILE	.005	.003	.001
42	ASN	.605	.609	.612	92	ALA	.013	.010	.004
43	VAL	.207	.225	.291	93	ASN	.022	.028	.019
44	ASP	.952	.951	.918	94	ILE	.006	.004	.003
45	GLY	.391	.397	.389	95	ASP	.022	.033	.025
46	SER	.368	.355	.338	96	GLY	.012	.007	.009
47	LEU	.024	.040	.043	97	THR	.003	.005	.004
48	LYS	.059	.083	.077	98	LEU	.010	.013	.007
49	TRP	.062	.063	.060	99	LYS	.218	.187	.271
50	GLN	.342	.358	.333	100	TYR	.017	.005	.019
51	PRO				101	GLY			

Figure S5: Selected curves from the diffusion measurements, 310K, D_2O

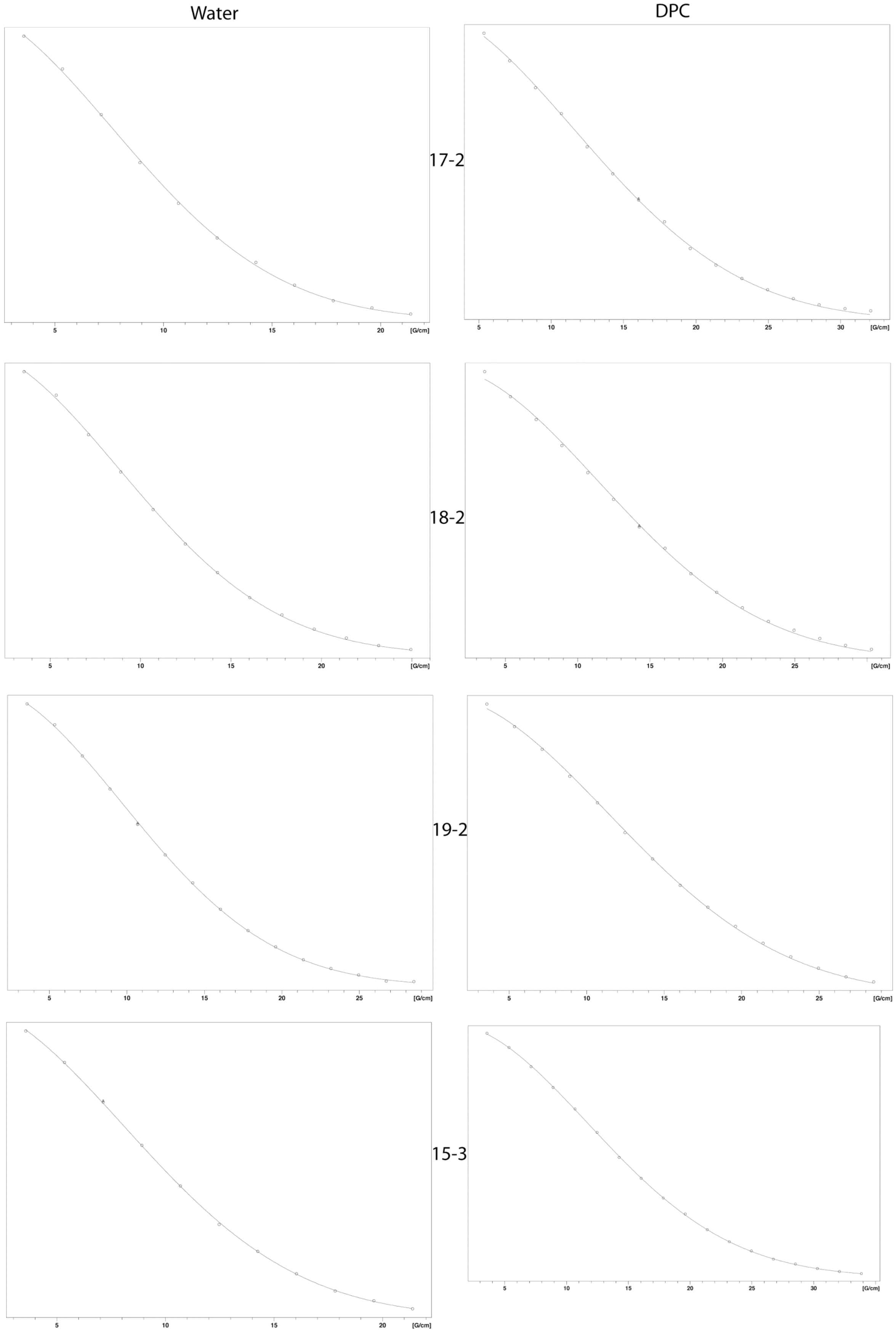


Fig. S6: Model of the glycolipid **17-2**, in which the attenuation of carbohydrate proton signals due to the paramagnetic relaxation with the micelle-integrating spinlabel 5-doxylstearate has been encoded into the size of the atom radii (large radii correspond to little attenuation, small radii to large attenuation). The lipid anchor extends at the bottom of the figure and has been omitted for clarity. The bond angle about the glycosidic bond linking the two carbohydrate moieties has been adjusted to result in a configuration compatible with the SL data. However, there is flexibility about this bond, and the precision of the SL data is clearly not sufficient to define the carbohydrate conformation more precisely.

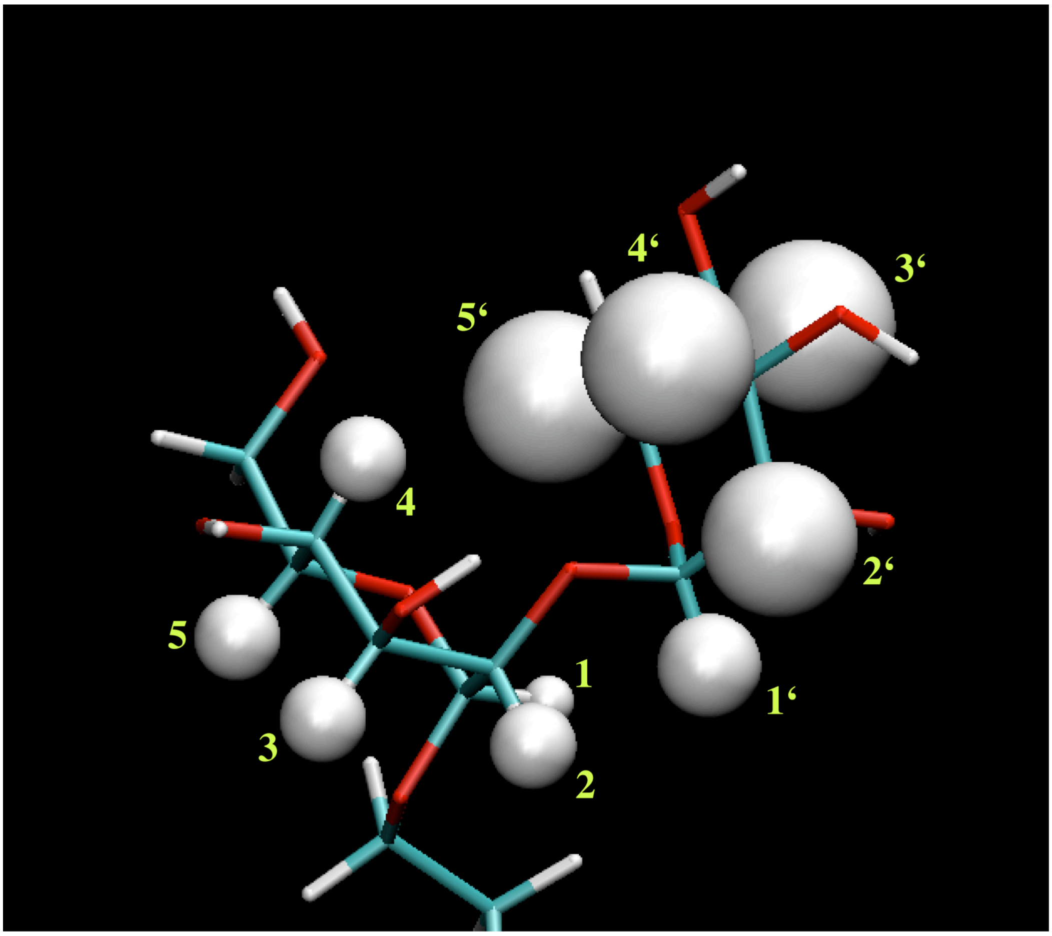


Table S5: Attenuations as extracted from i) the 1D proton spectrum, and ii) crosspeaks from a 80ms TOCSY spectrum

H Atom No.	**17-2**	**19-2**	
1D data	Signal(SL)/Signal(Ref)		
1	0.10	1	0.43
3	0.30	2	0.62
1’	0.22	1’	0.53
2’	0.51	2’	0.78
2D data			
1<->2	0.02		
2<->3	0.10		
3<->4	0.10		
3<->5	0.05		
1’<->2’	0.15		
2’<->3’	0.60		
3’<->4’	0.60		
4’<->5’	0.60		

Figure S7: Chemical shift changes upon adding **17-2** to CV-N color-coded onto the structure (left), as well as *differences* of chemical shift changes induced by adding **17-2** or **19-2**.

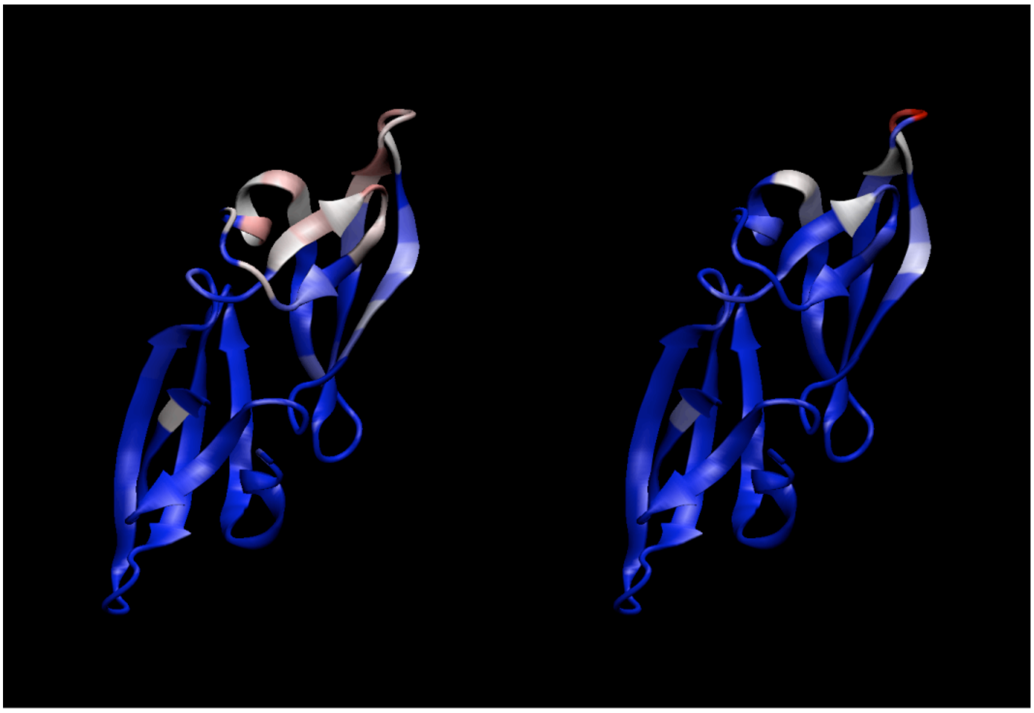
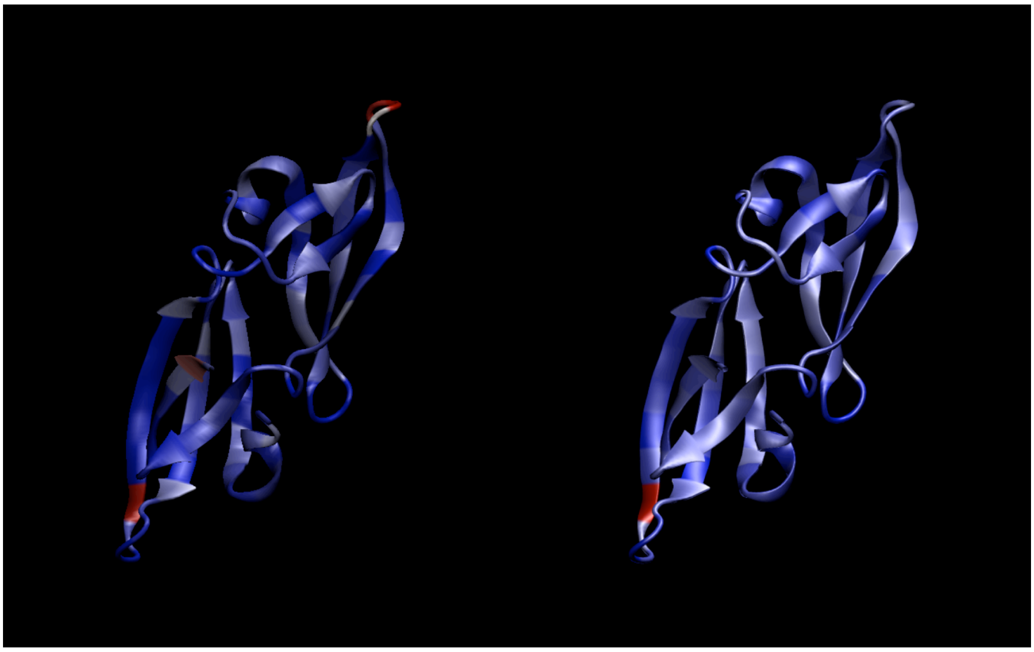


Figure S8: Attenuations of resonances of CV-N due to the presence of the spin label 5-doxylstearate upon adding **17-2** (left) or **19-2** (right) to CV-N color-coded onto the structure:



8. Interactions of Lipopolysaccharide and Polymyxin Studied by NMR spectroscopy

In the light of occurrence of bacterial strains with multiple resistances against most antibiotics, antimicrobial peptides that interact with the outer layer of gram-negative bacteria, such as polymyxin (PMX), have recently received increased attention. Here we present a study of the interactions of PMX-B, -E and -M with lipopolysaccharide (LPS) from a deep-rough mutant strain of *E. coli*. A method for efficient purification of biosynthetically produced LPS using RP-HPLC in combination with ternary solvent mixtures was developed. LPS was incorporated into a membrane model, dodecylphophocholine (DPC) micelles, and its interaction with polymyxins was studied by heteronuclear NMR spectroscopy. Data from chemical shift mapping using isotope labelled LPS or labelled polymyxin, as well as from isotope-filtered NOESY experiments, highlight the mode of interaction of LPS with polymyxins. Using MD calculations in explicit water the complex of LPS with PMX-B in the presence of DPC micelles was modelled using restraints derived from chemical shift mapping data and intermolecular NOEs. In the modelled complex the macrocycle of PMX is centered around the phosphate group at GlcN-B, and additional contacts from polar sidechains are formed to GlcN-A and Kdo-C, while hydrophobic sidechains penetrate the acyl chain region.

in revision: Mareš, J., Kumaran, S., Gobbo, M., Zerbe., O.
J. Biol Chem

Introduction

Cellular membranes segregate the interior of cells from their surroundings and therefore are crucial to maintain cells as autonomously functioning systems [1]. The chemical constituents of outer membranes from mammalian cells and bacteria are fundamentally different. The mammalian outer membranes are largely formed by phospholipid bilayers, whereas additional coating structures are present covering these in bacteria. In gram-positive bacteria a thick peptidoglycan layer is built around the phospholipid bilayer. In gram-negative bacteria, the peptidoglycan structure is much thinner and coated by an additional phospholipid-containing bilayer, whose outer leaflet is mainly composed of lipopolysaccharides (LPS) [2]. LPS are high-molecular-weight, strongly negatively charged molecules, that for smooth LPS can be divided in three regions: the lipid A portion of LPS inserts into the phospholipidic membrane and in many Gram-negative bacteria consists of a di-glucosamine diphosphate with 5 to 7 fatty acid chains extending to one side of the disaccharide. The lipid A is appended to a region (the inner core) of 8-12 variable sugars (including the negatively charged 3-deoxy-D-manno-oct-2-ulosonate (Kdo) units) and 3-8 phosphate residues. To the inner core is covalently associated the O-antigen, an oligosaccharide chain of variable length and chemical composition, depending on the exact type of LPS.

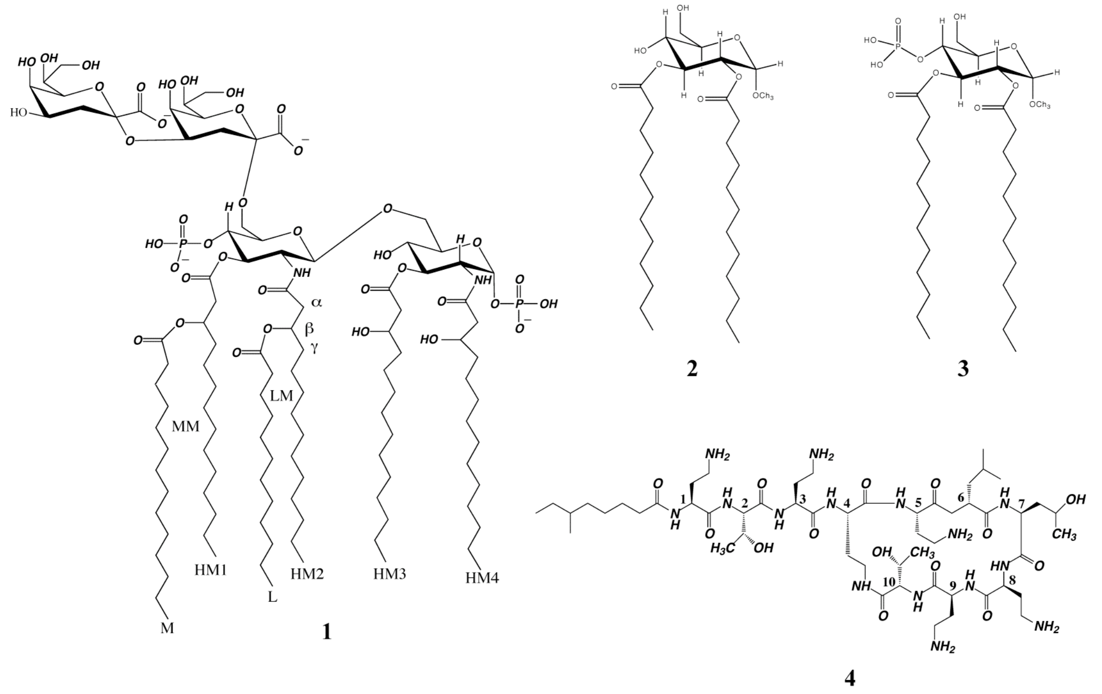


Figure 1: Chemical structures of LPS from the D31m4 *E. coli* strain (**1**), the two model compounds for LPS (**2** and **3**) and PMX-M (**4**). Note that in PMX-B residue D-Leu-6 is replaced by D-Phe and Thr-7 is replaced by Leu, and in polymyxin-E Thr-7 is replaced by Leu. In the text, the glycosidic residues of LPS are identified with the letters A-D, starting from the far right GlcN residue.

Antimicrobial peptides against gram-negative bacteria can interfere with the integrity of this LPS layer. Even if no specific interaction with LPS exists, they need to pass the LPS layer in order to gain access to the cytoplasmatic membrane. Cationic lipopeptide antibiotics like polymyxin B and E, characterized by a heptapeptide ring and a fatty acid tail (Fig. 1), are used to treat infections with gram-negative bacteria. However, their severe toxic side effects have limited their usage to treatments against bacteria resistant against most other antibiotics such as *Pseudomonas aeruginosa*. The interaction of LPS with various antimicrobial peptides has been the subject of a number of studies [3-11]. In some of these, the conformation of the LPS-bound peptides was established using transferred NOE effects, and the complex between LPS and the peptides was established by docking the trNOE-derived peptide conformer to LPS [7], whose coordinates were taken from the crystal-structure of FhuA-bound LPS [12].

The emphasis of this work was to obtain experimental data on the LPS-PMX complex, thereby allowing a detailed understanding of the interacting moieties. Since LPS in the outer membrane of *E. coli* cells is integrated into a phospholipid bilayer, it was studied while integrated into phospholipid micelles to better mimic the natural environment. In our studies we used LPS from the deep rough mutant D31m4 of *E.coli*. Biosynthetic production of the latter and its isolation and purification from the membrane-fraction of the corresponding cells was described in literature. However, to facilitate purification by RP-HPLC, the phosphate groups were methylated[13]. The modified LPS was investigated in detail by NMR [14]. Since interactions with the charged phosphate groups were proposed to be important for binding antimicrobial peptides, we developed an HPLC-based method that allows purification of LPS in its natural (non-methylated) form (Fig. 1). ^{13}C,^{15}N-Labeled LPS from the deep rough mutant strain was isolated and purified to chemical homogeneity. Extensive use of heteronuclear solution NMR techniques allowed characterization of LPS embedded in DPC micelles, and facilitated the study of its interactions with polymyxins from different organisms. The interaction studies relied on chemical shift mapping techniques and isotope-filtered NOEs, and allowed direct probing for the interaction sites.

8.1 Experimental Procedures

Materials $^{15}NH_4Cl$ was purchased from Spectra Isotopes (Columbia, USA), perdeuterated DPC-d38 (99%-d), and D_2O were ordered from Cambridge Isotope Laboratories (Andover, Massachusetts, USA). Methyl-5-doxylstearic acid was bought from Aldrich (Buchs, Switzerland). The Re-LPS producing strain D31m4 was purchased from the *E. coli* Genetic Resource Center, New Haven. The PMX-M producing strain *Paenibacillus kobensis* M was obtained from Prof. J.C. Vederas and PMX-B and -E were purchased from Sigma Aldrich Laboratories.

Production of ^{13}C-labelled LPS from the E. coli strain D31m4 Cells from the D31m4 strain of *E. coli* were grown at 37°C to an OD of around 1.0 at 600 nm on minimal medium M9 using 4 g ^{13}C glucose and 1 g ^{15}N ammonium chloride supplemented with 100 mg of Trp, His and Pro per liter.

After harvest cells were resuspended in 50 ml of ice-cold water and pelleted down. To the pellet a minimum amount of cold water was added such that a thick paste was formed. LPS together with other components was precipitated through addition of 90 ml ice-cold methanol and centrifuged at 8000 g. The pellet was resuspended in 90 ml ice-cold acetone, homogenized and centrifuged again, followed by another acetone washing step. The lyophilized cells (ca. 0.7 g) were taken up in 50 ml of a phenol:chloroform:petroleum ether (4:10:16; v:v) solvent mixture and centrifuged at 9200 g, after which most of the LPS was contained in the supernatant. The remaining pellet was extracted once more to increase the yield in LPS. The supernatant was concentrated under a nitrogen stream and 2 ml of water were added dropwise to the concentrate. A waxy precipitate was formed followed by three cycles of washing with methanol and subsequent centrifugation. Thereafter the pellet was dried and lyophilized, after which it could only be resuspended/dissolved in water using repetitive additions of small amounts of water followed by sonication. Solubilization was improved upon adding aqueous 0.1M EDTA in the first portions. The resulting solution was centrifuged at 200000 g overnight and then lyophilized. The lyophilized pellet after resuspension in mixture of solvent A (methanol:chloroform:water 57:12:31 v/v/v), and aqueous EDTA was directly loaded onto the HPLC column.

For optimal purification of LPS a gradient system involving ternary solvent mixtures was used consisting of solvent A in 10mM NH_4Cl and methanol:chloroform (29.8:70.2 v/v, solvent B) in 50mM NH_4Cl. LPS (6 mg) was dispersed in a two phase system formed from 0.8 ml of solvent A and 0.2 ml of 0.1 M aqueous EDTA pH=7 and loaded directly onto the RP-C8 column. Chromatographic separation was achieved using the following gradient of solvents A and B: 2 CV of 2% B, 3 CV (2-17% B), 3.5 CV (17-27% B). UV detection was impossible and hence fractions were lyophilized and their content checked by MALDI-TOF using 6-aza-2-thiothymine as the matrix. Elution of the desired LPS occurred around 20-23% of solvent B.

Production of ^{13}C,^{15}N-labelled PMX-M from P. kobensis M The producer strain, *Paenibacillus kobensis* M, was grown aerobically at 30 °C on tryptic soy agar (TSA). A 1-liter batch of M9 medium was inoculated with a 10-ml *P. kobensis* M preculture (1% inoculum). After a total growth time of 16–24 h at 30°C with shaking (200 rpm), the cells were removed by centrifugation (1 hr, 10,000 rpm) and the supernatant was then passed through a Amberlite XAD-16 column. After washing with 30% ethanol, active peptide was then eluted with 70% isopropanol, which was adjusted to pH 2 with 12N HCl.

All fractions were assessed for antimicrobial activity using a well plate assay. The contents of the active fraction were applied to a Superdex peptide 10/300 column (Amersham Biosciences). Fractions were collected for 3 column volumes with pure MilliQ Water and each assayed for activity. All active fractions were pooled, concentrated and applied as 20% isopropanol solutions to C18 reverse-phase HPLC. Complete purification required two separate steps of C18-HPLC. The first separation used a gradient of water/isopropanol gradient (0.1% trifluoroacetic acid), from 20% to 50% isopropanol, and the second step a water/methanol gradient (0.1% trifluoroacetic acid), from 45% to 85% methanol. PMX-M eluted at around 55%. Finally, 8 - 10 mg of pure PMX-M was obtained as slightly yellowish powder from a 1-liter culture and its chemical nature verified by ES1-MS (exp. mass: 1224.73Da, theoret. mass:1223.57 Da).

During all steps of expression and purification antimicrobial activity was monitored by inhibition of growth of an indicator strain. Agar plates were prepared by inoculating molten TSA (40 g/liter) with a culture of the indicator organism *E. coli* (1.0% inoculum). Small wells (app 4.6 mm diameter) were made in the seeded agar plates and 50-μl of filtered culture supernatant were added to the wells. Plates were incubated at 30°C, and the growth of the indicator organism was visible after approx. 3 h. [15]

NMR spectroscopy LPS samples used in this work contained approx. 1mM LPS, 300 mM d_{38}-DPC in 40 mM d_{13}-MES D_2O buffer pD=5.8. Interaction studies with PMX were performed in acetate buffer in D_2O or H_2O/D_2O 9/1 pH=4.4. All spectra were recorded on Bruker AV-600 or AV-700 NMR spectrometer at T=310K. Proton and carbon chemical shifts were calibrated to DSS, and nitrogen shifts were referenced indirectly to liquid NH_3 [16]. The spectra were processed using the Bruker Topspin2.0 software and transferred into CARA [17] or SPARKY[18] programs for further analysis.

For chemical shift assignments of $^{13}C,^{15}N$-labeled LPS 2D versions of 3D double- and triple-resonance experiments were recorded. In general, experiments used coherence selection schemes via pulsed-field gradients [17] and sensitivity-enhancement building blocks [19, 20] whenever possible. For assignments of the carbon spin systems in the lipid chains and the sugar units (H)CCH experiments recorded with 4 and 12 ms DIPSI-2 C-C mixing cycles were used. Linkage of the lipid chains onto the glucosamine parts of lipid A was achieved via correlations with the amide nitrogens using HNCA and HN(CO)CA experiments. To distinguish the two Kdo units key NOEs derived from a ^{13}C-resolved NOESY were exploited. Assignment of all resonances of polymyxin was done using HN(CO)CACB, HNCACB [21] and (H)CCH experiments [22, 23] analogous to the procedure used for proteins. Because of the small size of the peptide 2D versions were recorded with a total of less than 12 h measuring time for acquiring all spectra. Assignments of polymyxin-B and –E were based on assignments from PMX-M adjusted by using additional 2D heteronuclear spectra.

In the spin label experiments, a 0.5 mM solution of LPS was separated into two aliquots, and to one of these 5-doxylstearate methyl ester was added so that the final concentration corresponded to approximately one spin-label per micelle. Signal intensities from the two corresponding constant-time [$^{13}C,^{1}H$]-HSQC were extracted and the ratio of signal intensities from the samples with and without spin label was calculated.

Measurements of interactions between LPS and PMX by chemical shift mapping observing LPS resonances utilized a 350 μM solution of $^{15}N,^{13}C$ Re-LPS and equimolar unlabelled peptides. Chemical shift changes in PMX-M were monitored in 200 μM solutions of $^{15}N,^{13}C$ -labeled PMX-M and equimolar unlabelled Re-LPS.

Molecular dynamics calculations All calculations were preformed within the program GROMACS [24]. Briefly, coordinates of LPS were adapted from the pdb entry 1QFF and coordinates of polymyxin B were built using the program Ghemical

[25]. Parameters and topologies of PMX-B and LPS as well as partial charges of PMX-B for GROMACS were established based on data from PRODRG server[26] and the GROMOS 53a6 forcefield [27]. Partial charges of LPS were assigned for a protonation state corresponding to experimental conditions of pH by the MPEOP method [28, 29]. Parameters for DPC were derived from values for DPPC from the GROMOS force field library. A detailed description of the methodology persued can be found in the Supp. Mat. Briefly, the system was prepared as follows: i) An initial complex between LPS and polymyxin B was prepared and equilibrated in the presence of a DPC micelle. ii) A set of simulated annealing calculations was performed yielding 450 structures. The DPC molecules were explicitly included in the system, water molecules were substituted by implicit solvent and the dominant NOE-derived upper distance limits were included (force constant was set to 1000 [kJ mol^{-1} nm^{-2}]). iii) Those structures were selected, which displayed best agreement with the chemical shift mapping data. iv) These were then equilibrated with explicit solvent and subjected to further refinement and analysis. The latter included an assesment of the stability of the MD trajectory and a comparison of the average intermolecular distances with chemical shift mapping data.

8.2 Results

Production and purification of ^{13}C-labeled LPS *E. coli* strains from the deep rough *E. coli* mutant D31m4 were isolated from the membrane fraction of cells grown on minimal medium containing ^{13}C-glucose and ^{15}N-NH_4Cl as the sole carbon and nitrogen sources, respectively. After PCP extraction and further purification using published protocols [30] the yield was ca. 129 mg/l of culture. Remaining impurities were removed by RP-HPLC using a ternary solvent mixture. In this procedure the solvent system was carefully adapted to form a single phase over the whole gradient range of solvent A (methanol:chloroform:water) and solvent B (methanol:chloroform) at RT. Importantly, any mixture of these two solvent systems is relatively close to a two-phase system, and this condition proved to have favorable properties for dissolving LPS. MALDI-TOF spectra of LPS before and after this HPLC purification step are depicted in Fig. 2. Sufficient quantities (40 mg) of chemically pure LPS for the NMR studies could be produced from 1 L of culture using this protocol. As demonstrated in Fig. S4 in the Supp. Mat. this method is capable of separating pyrophosphate from the monophosphate derivatives.

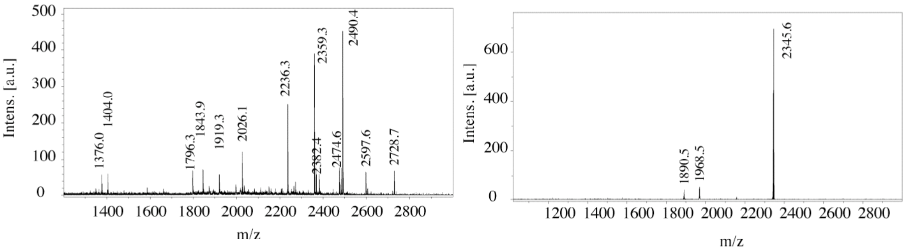


Figure 2: Comparison of the MALDI-TOFF MS spectrum of commercial LPS with LPS from the deep rough D31m4 *E. coli* strain purified by the protocol that includes an additional HPLC purification step.

Assignment of LPS and polymyxin resonances Chemical shift mapping [31] or NOE-based methods [32] can be used to study biomolecular interactions (see also [33-35]). Both methods potentially deliver information on interacting moieties but require assignments of chemical shifts. The best chemical shift dispersion is usually available in heteronuclear shift correlation spectra (e.g. [^{15}N,^{1}H]-HSQC or [^{13}C,^{1}H]-HSQC spectra). Importantly, these experiments still work well in the presence of the increased line widths that are usually present in systems that are stably anchored into phospholipids micelles. In addition, as was unfortunately the case in some of our applications, additional exchange broadening occurred upon complex formation. To probe integration of LPS into DPC micelles and to study its interaction with peptides we decided to label it with ^{13}C and ^{15}N isotopes and use the corresponding HSQC spectra for chemical shift mapping.

To assign all signals in the constant-time [^{13}C,^{1}H]-HSQC spectrum [36], 3D (H)CCH-TOCSY spectra [22, 23] served for assignment of spin systems (Fig. 3). The C,H-plane of the HNCA [19, 37] and HN(CO)CA [37] experiments were used to establish scalar connectivities between terminal carbons of the fatty acid chains and C-2 of the glucosamine moieties. Hydroxymyristoyl (HM) can be distinguished from lauroxymyristoyl (LM) and thereby helps to differentiate between GlcN-A from GlcN-B (see legend of Fig. 1). Due to chemical shift degeneracies it was impossible to assign chains of the fatty acids (myristoyl of myristoxy-myristoyl and lauroyl of lauroxymyristoyl). The two Kdo units were linked and thereby distinguished from each other using several key NOEs in the 3D ^{13}C-NOESY spectra. The unique chemical shifts of the C3 moiety of Kdo is located in a region separated from all other sugar resonances, and was used as a starting point for sequential assignment. Only the methylene group C3 of Kdo-C is expected to receive an NOE from H(C6) of GlcN-B. This assignment was additionally supported by the fact that H(C6) and H(C7) of Kdo-C displayed an NOE to H(C3) of both Kdo units, which is unlikely to be the case for H(C6) of Kdo-D. The ^{1}H,^{13}C and ^{15}N chemical shifts of LPS in DPC micelles are reported in Table 1.

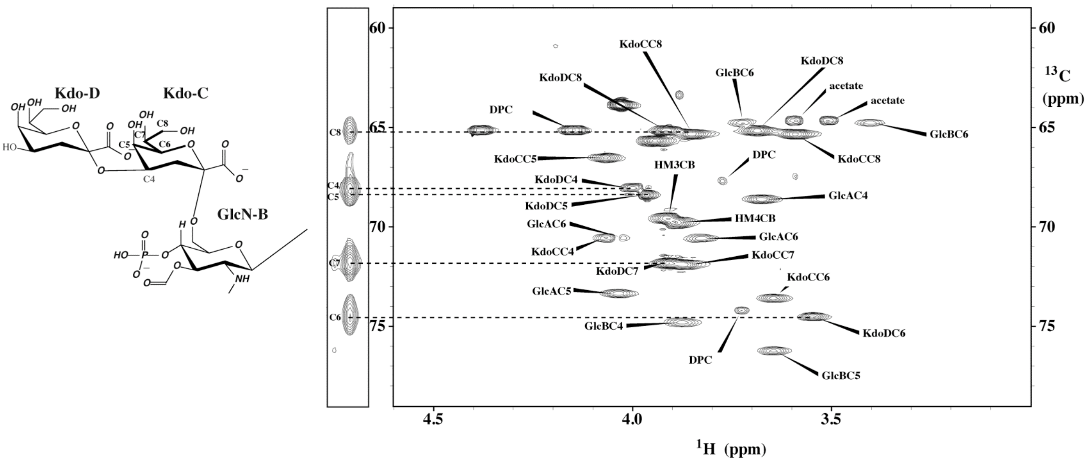


Figure 3: Assignment of ^{13}C spin systems of a short stretch contained in Kdo-D from LPS using (H)CCH-TOCSY spectra. On the right the constant-time [^{13}C,^{1}H]-HSQC spectrum is shown, in the middle a strip from the (H)CCH TOCSY taking at the proton frequency of H-(C4) of Kdo-D.

Topology of LPS and polymyxin insertion into the DPC micelle To probe whether LPS properly inserts into the DPC micelles, and whether the sugar moieties really protrude into the aqueous phase, the micelle-integrating spin label methyl 5-doxylstearate was used. The paramagnetic moiety of the spin label resides in the headgroup region [38, 39], and signal reductions in the proton-carbon correlation maps indicate proximity to the water-micelle interface. The results from the spin-label experiment are depicted in Fig. 4. Strongest attenuations of signals are observed for C,H moieties of terminal acyl chain carbons, that are close to the GlcN sugars. Much less pronounced attenuations are observed for signals from the GlcN moieties, and the Kdo units are essentially not affected. The data demonstrate that LPS inserts into the DPC micelles such that the amide moieties of LPS are located in the headgroup region, the acyl chains are inserted into the micelle interior, and the carbohydrate units are exposed on the surface of the micelles.

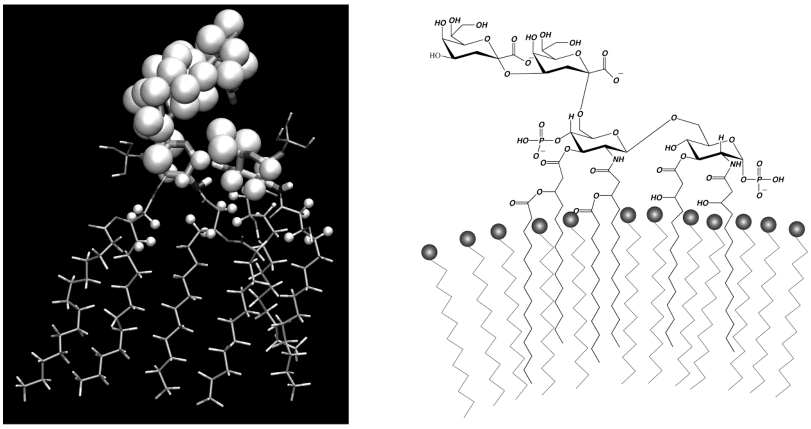


Figure 4: Left: Results from the spin-label experiments displayed on the structure of LPS. The size of the sphere is proportional to the remaining signal in the proton-carbon correlation map. Because signals from ends of the fatty acid chains are not resolved in the spectra, no encoding is shown for these atoms. Right: Sketch of LPS from the D31m4 strain when inserted into DPC micelles. The phospholipid headgroups are depicted as grey spheres.

We additionally probed for the topology of the polymyxin-micelle complex using the micelle-integrating spin-label methyl 5-doxylstearate as described in the previous section. Notably PMX-B, -E and -M differ only for residues in position 6 and 7 (see legend of Fig.1). Largest attenuations were observed for the amide moieties of the first residue, and for residues 6 and 7, the latter two present hydrophobic or aromatic residues (Table S11), revealing that polymyxins are attached via the N-terminal lipid chain and the apolar residues to the micelle.

Interactions of LPS with various types of polymyxins Studies of interactions of LPS with polymyxin were performed using chemical shift mapping and isotope-filtered NOESY experiments. We initially used very simplified model compounds of LPS derived from α-(D)-glucose, that lack Kdo and the second glucosamine units (Fig.1). Considering the proposed importance of phosphate moieties for binding it was not surprising that compound **3** did not display any interactions with PMX-B, but also compound **4**, that contains the presumably important phosphate group, did not cause any changes in the spectra, indicating that the Kd is (much) larger than 1mM.

The [^{15}N,^{1}H]- and the [^{13}C,^{1}H]-HSQC spectra of PMX-M revealed large changes (*vide infra*). The chemical shift changes of LPS carbon and proton frequencies upon adding PMX-B,-E or -M are mapped onto the structure in Fig. 5. In general, the largest chemical shift changes of LPS resonances were observed for all observable atoms of GlcN-B. Interestingly, resonances from C1 and C2 of GlcN-A were affected only very little (Fig. S3A and S3C). In addition, resonances from Kdo-C display significantly larger changes than those of Kdo-D, with C-3 of Kdo-C being shifted by the largest extent (see Fig. S3F). Moreover, large changes in chemical shift are observed for the alpha (Fig. S3G), beta and gamma (Fig. S3H) positions of the fatty acid chains, in particular for those of HM2 (note that the second branch of MM and LM cannot be assigned due to spectral overlap). The Cα change for HM1 is less pronounced. The alpha position of HM4 was apparently more affected than of HM3. The chemical shifts changes on the beta position were smaller and clear only for HM4.

A comparison of chemical shift changes of LPS between its complexes with PMX-B, PMX-E and PMX-M may serve to identify differences in their binding modes. Chemical shift changes are almost identical at positions C-4, C-5 of GlcN-B as well as on position C-3 of Kdo-C for PMX-B and PMX-E (Fig. S3E and S3F), and very similar changes occur for the alpha position of HM2. We suspect that these regions are in contact with the conserved parts of polymyxins. However, significant differences were observed for C-3 of GlcN-B and moderate differences can be seen at positions of C-3, C-5 and C-6 of GlcN-A and at position C-1 of GlcN-B (Fig. S3B). Addition of PMX-M results in line-broadening for resonances of C-4 and C-5 of GlcN-B, as well as for C-3 of Kdo-C, and in a chemical shift difference for C-6 of Kdo-C. We address this observation to a slightly different binding mode of PMX-M, which results in different contacts to Kdo-C.

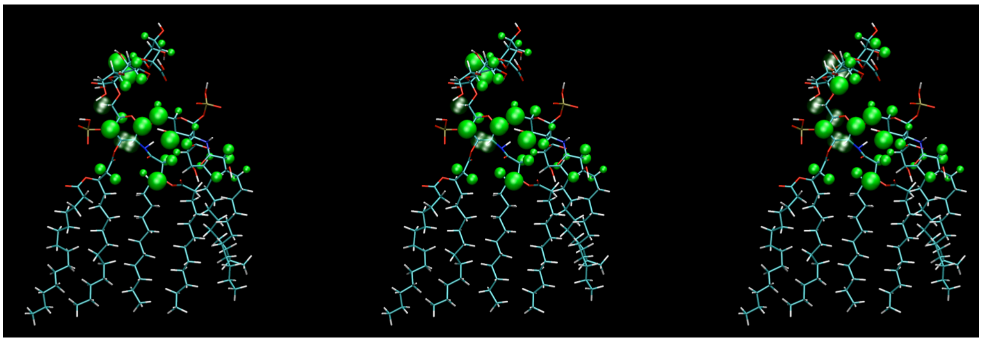


Figure 5: Changes in ^{1}H and ^{13}C chemical shifts of resonances from LPS when adding PMX-B (A), PMX-E(B) and PMX-M(C). The size of the spheres is proportional to the observed differences. Deviations of peak positions were extracted from the [^{13}C,^{1}H]-HSQC spectra and computed according to $\Delta\delta=\mathrm{sqrt}(((\Delta\delta\ C)^2+(10^*\ \Delta\delta\ {}^1H)^2$. Those atoms that are exchange-broadened beyond detection are depicted by green transparent spheres.

Interactions between LPS and polymyxin were additionally detected in heteronuclear spectra when ^{13}C,^{15}N-labelled PMX-M was mixed with unlabeled LPS. Unfortunately, very broad lines and multiple sets of resonances were detected in the [^{15}N,^{1}H]-HSQC of polymyxin-M in presence of LPS indicating the presence of exchange processes. Moreover, the fact that the signals were shifted considerably and, even more importantly, the presence of very broad line precludes detailed assignment. Even when changing pH or adding other detergents such as SDS we have not been able to achieve conditions that result in spectra of sufficient quality. Considering that interactions between LPS and polymyxin are suggested to be electrostatic in nature [40], we decided to reduce the strength of the interaction by adding small amounts of divalent metal ions. When adding PMX-B, -M or –E in the presence of Ca^{2+} or Mg^{2+} we noticed slight precipitation. In the presence of 20 mM $MgCl_2$ a good-quality [^{15}N,^{1}H]-HSQC was obtained, and the peak positions are close to those in the absence of LPS. Fortunately, linewidths in the [^{13}C,^{1}H]-HSQC spectra are smaller such that interactions can be directly detected by shift mapping methods. The results are summarized in Table S3, and shall only be briefly summarized here: largest chemical shift changes occur for Cα resonances of residues 4, 5, 6, 8 and 9 of polymyxin-M, the members of the macrocyclic ring. Interestingly, smaller differences were observed for resonances from the lipid chain or from Dab-1, Thr-2 or Dab-3. The fact that only residues from the macrocycle experience large changes, and the observation, that the changes in the macrocycle occur at certain but not in all positions, indicates that the heptapeptide ring of PMX binds in a rather well-defined manner to LPS.

In addition to chemical shift mapping we have also conducted isotope-filtered NOESY experiments, in which ^{13}C-labeled LPS was mixed with unlabelled PMX-B, -E, so that the spectra only contain intermolecular NOEs (see Table 2). Interestingly, in the case of PMX-B, NOEs between side chain protons of Phe-6 or Leu-7 and the lipid chains of LPS are detected, supporting the results from the shift mapping data. In particular, a number of NOEs between protons of the π-system of Phe-6 and the lipid chains are observed, in particular to Cγ and Cα of HM3. In the case of PMX-M position 7 is occupied by the polar Thr residue. The fact that PMX-B or PMX-E form more contacts with LPS that involve side chains of residues 6 and 7 indicates that the hydrophobic nature of this side chain may be important for orienting polymyxin in the complex with LPS.

To obtain a first picture of the complex formed between LPS and PMX-B we have performed a MD calculation in which LPS was docked to PMX-B using NOE-derived upper distance limits (Fig. 6) in explicit water and in the presence of a DPC micelle. In the complex backbone amide moieties of the peptide macro-cycle form contacts with the phosphate group of GlcN-B. Residues from one hydrophobic site - D-Phe-6 and Leu-7 - are in proximity to the fatty acid chains of LPS. Leu-7 makes additional contacts with the GlcN-B moiety. The amino group of Dab-8 is involved in electrostatic contacts with the phosphate group at GlcN-A. Amino groups of Dab-9 and -3 are possibly forming electrostatic interactions with the carboxyl group of Kdo-C. The backbone of residues Dab-1 to Dab-3 is located in the acyl region of the branched fatty acids originating from GlcN-B. To summarize most contacts in this model structure are made with GlcN-B and the adjacent atoms of GlcN-A, as well as with parts of Kdo-C, and the model structure is supported by the chemical shift mapping data. In addition, interacting moieties of LPS that make contacts common with all polymyxin variants are in contact with conserved parts of these peptides (residues 1 to 4 and 8 to 10), while those sites, that differ in their interaction with the different peptides form contacts with non conserved residues.

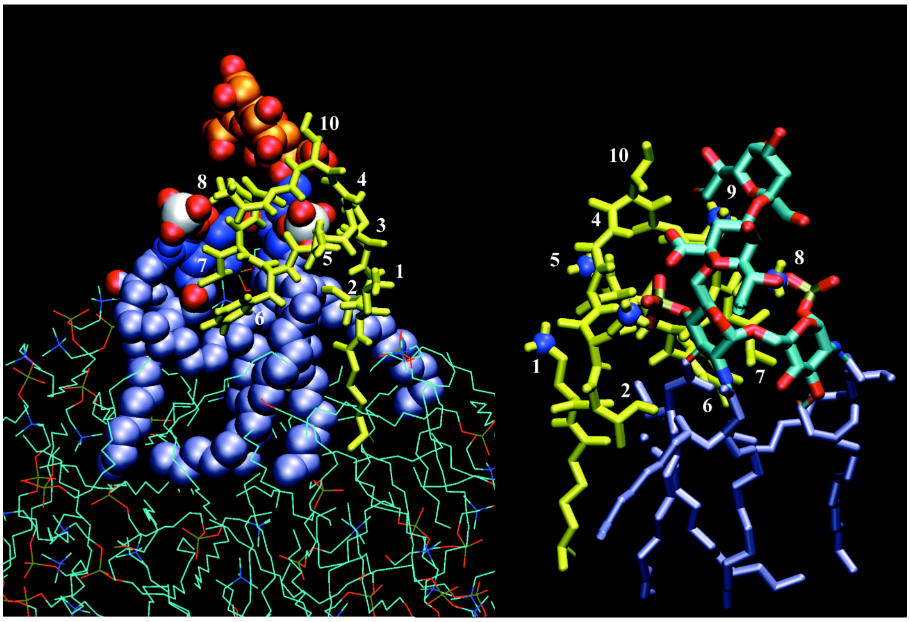


Figure 6: Left: Structure of the PMX-B—LPS complex as derived from the MD calculation. The covalent structure of PMX-B is depicted in yellow. Heavy atoms of LPS are drawn as van der Waals spheres, with lipid chains colored in ice-blue, GlcN carbons in dark blue, Kdo carbons in orange, and phosphorous atoms in white. DPC molecules are indicated by thin lines. Residue numbers of PMX-B are placed close to the corresponding sidechains. Right: Representation showing only bonds of LPS and PMX-B. PMX-B is again drawn in yellow, with amino nitrogens of DAB residues in blue, while LPS bonds are drawn in ice-blue (lipids) or green (sugar parts).

8.3 Discussion

Sepsis caused by gram-negative bacteria is a serious source of mortality in many clinical cases, accounting for approximately 200,000 deaths in the US annually (see David [10] and references therein). The primary trigger for sepsis was identified as endotoxin, and endotoxin-neutralizing agents are therefore valuable therapeutics. PMX-B is considered as the “gold-standard” for LPS-sequestering agents. It is a highly cationic decapeptide containing six diaminobutyric acid (Dab) residues, a macrocylic ring involving residues 4 to 10, and an acyl chain coupled to the N terminus. In this work we have used polymyxins of different types to study interactions with LPS from the deep rough mutant *E. coli* strain. To our knowledge this work for the first time presents experimental data on the interacting moieties in a membrane mimetic environment.

The conformation of LPS-bound PMX-B and -E was elucidated previously using transfer-NOE techniques by Pristovsek. [7] Therein, the LPS-bound peptides assume envelope-like bent cycles that separate the two hydrophobic residues 6 and 7 from the charged Dab residues 4, 5, 8 and 9. The conformation from the transfer-NOE derived structure was then docked onto a lipid A model. In the resulting complex the lipid chain forms transient contacts with acyl chains A and B from lipid A. Moreover, the two phosphate groups are in close contact with the γ-NH_2 groups of Dab residues 1/5 and 8/9. The data highlight the importance of electrostatic interactions between phosphate groups and amino groups of polymyxin as well as hydrophobic interactions of the lipid chains with Phe or Leu residues. Martin *et al.* have studied the structure of

LPS-bound PMX-M [15], which displayed a chair-like confor-mation, in which the side chain of Dab-4 or -8 and Leu-6 or Thr-7 point into the opposite directions in a fashion similar to that proposed for PMX-B and –E [15]. Pristovsek determined the structure of a LPS-bound synthetic fragment of the LALF protein [6] revealing a hairpin-type fold again characterized by spatial separation of hydrophobic and cationic residues. In another study Bhunia *et al.* investigated binding of melittin to LPS micelles [41]. Cationic residues have also been postulated to critically contribute to binding of LPS to proteins such as MD-2 [42] or FhuA [12, 43].

Herein we set out to determine interactions between LPS and polymyxins in more detail. Considering that interacting chemical moieties have not been experimentally identified with confidence so far we reasoned that these studies should use chemically defined environments in a setup that mimics natural conditions. To achieve chemical homogeneity we used biosynthetic LPS to facilitate isotope labeling. LPS from the deep rough mutant of *E. coli* presents a simple system containing all chemical moieties believed to be important for the interaction. LPS in gram-negative bacteria is embedded in a phospholipid bilayer, and therefore the system was studied in DPC micelles. It could be shown that LPS properly inserts into the micelle via integration of the acyl chains into the micelle interior, and the GlcN and Kdo portions are fully exposed in the aqueous compartment. LPS from the same *E. coli* strain has been assigned previously using homonuclear 2D NMR techniques.[13, 14].

For the purpose of the study the development of chromatographic techniques for LPS was very important that allow studying the unmodified LPS. Using ^{13}C isotope labeling we have been able to fully assign LPS while integrated into DPC micelles. In the latter environment spectra with reasonable line widths and resolution can be recorded, allowing the study of interactions with polymyxin in sufficient detail. Our experimental data now confirm the model of the PMX-B LPS complex, proposed by Pristosek et al. [7] In their complex the macrocycle of PMX-B covers the GlcN disaccharide unit, and hydrophobic sidechains of polymyxin form contacts with the lipid chains from LPS, and a similar complex topology was proposed for the interaction of LPS with PMX-M [15]. Such an arrangement requires a reasonable separation of hydrophobic and hydrophilic side chains in polymyxin, so that hydrophobicity matching between polymyxin and LPS can help to orient it. It should be noticed that LPS in its natural environment is embedded in a phospholipid environment. Hydrophobicity matching of moieties of polymyxin with the corresponding parts of LPS in that environment is amplified, because similar requirements influence interactions of polymyxin with the phospholipids. It is therefore of little surprise that Leu or Phe residues, for which favorable energies for partitioning into the membrane interior or the water-membrane interfacial region have been measured [44], form contacts with the α, β, and γ acyl chain carbons of LPS. In PMX-M position 7 is occupied by Thr, a much more polar residue. According to the data from this work, PMX-M may be slightly differently oriented, possibly forming some-what stronger contacts to the Kdo units. In the PMX-M—LPS complex electrostatic or polar interactions involving the Dab residues of polymyxin and the carbohydrate moieties of GlcN-B or Kdo-C and to a smaller extent, Kdo-D dominate, while in the case of PMX–B or –E additional hydrophobic interactions from sidechains of Leu-7 are likely to contribute to binding. While our work confirmed the contacts made with moieties of the GlcN units proposed in the model complex from

Pristovsek [7] it is different in that it emphasizes the importance of contacts to Kdo units, in particular to Kdo-C.

Different models have been proposed for the mechanism of action of polymyxin. Shai et al. have proposed that binding of polymyxin to the core part of LPS, lipid A, results in disturbance of the LPS-phospholipid bilayer destroying the integrity of the outer membrane, and possibly leading to pore formation [3, 8]. The orientation of LPS in the phospholipid micelles as measured using spin labels has demonstrated that the carbonyl group region of the LPS lipid chains is located in the headgroup region. Binding of polymyxin to LPS therefore places the amphiphilic peptide in a similar position compared to binding to pure phospholipid micelles. Accordingly, similar mechanisms for membrane permeabilization are plausible.

ACKNOWLEDGEMENTS

We are very grateful to John C. Vederas for supplying us with cells producing polymyxin M and John Robinson for helpful discussions. We would like to thank for financial support from the Swiss National Science Foundation (grant No. 3100A0-11173).

References:

1. Gennis, R. B. (1989) *Biomembranes: Molecular Structure and Function*, 1.st edn, Springer, New York.

2. Raetz, C. R., Garrett, T. A., Reynolds, C. M., Shaw, W. A., Moore, J. D., Smith, D. C., Ribeiro, A. A., Murphy, R. C., Ulevitch, R. J., Fearns, C., Reichart, D., Glass, C. K., Benner, C., Subramaniam, S., Harkewicz, R., Bowers-Gentry, R. C., Buczynski, M. W., Cooper, J. A., Deems, R. A. & Dennis, E. A. (2006) Kdo2-Lipid A of Escherichia coli, a defined endotoxin that activates macrophages via TLR-4, *Journal of lipid research. 47*, 1097-111.

3. Rosenfeld, Y., Papo, N. & Shai, Y. (2006) Endotoxin (lipopolysaccharide) neutralization by innate immunity host-defense peptides - Peptide properties and plausible modes of action, *J Biol Chem. 281*, 1636-1643.

4. Rosenfeld, Y. & Shai, Y. (2006) Lipopolysaccharide (Endotoxin)-host defense antibacterial peptides interactions: Role in bacterial resistance and prevention of sepsis, *Bba-biomembranes. 1758*, 1513-1522.

5. Rana, F. R. & Blazyk, J. (1991) Interactions between the antimicrobial peptide, magainin 2, and Salmonella typhimurium lipopolysaccharides, *FEBS Lett. 293*, 11-5.

6. Pristovsek, P., Feher, K., Szilagyi, L. & Kidric, J. (2005) Structure of a synthetic fragment of the LALF protein when bound to lipopolysaccharide, *J Med Chem. 48*, 1666-1670.

7. Pristovsek, P. & Kidric, J. (1999) Solution structure of polymyxins B and E and effect of binding to lipopolysaccharide: An MMR and molecular modeling study, *J Med Chem. 42*, 4604-4613.

8. Papo, N. & Shai, Y. (2005) A molecular mechanism for lipopolysaccharide protection of gram-negative bacteria from antimicrobial peptides, *J Biol Chem. 280*, 10378-10387.

9. Ding, L., Yang, L., Weiss, T., Waring, A., Lehrer, R. & Huang, H. (2003) Interaction of antimicrobial peptides with lipopolysaccharides, *Biochemistry. 42*, 12251-12259.

10. David, S. (2001) Towards a rational development of anti-endotoxin agents: novel approaches to sequestration of bacterial endotoxins with small molecules, *J. Mol. Recognit. 14*, 370-387.

11. Andra, J., Lohner, K., Blondelle, S. E., Jerala, R., Moriyon, I., Koch, M. H., Garidel, P. & Brandenburg, K. (2005) Enhancement of endotoxin neutralization by coupling of a C12-alkyl chain to a lactoferricin-derived peptide, *Biochem J. 385*, 135-43.

12. Ferguson, A., Hofmann, E., Coulton, J., Diederichs, K. & Welte, W. (1998) Siderophore-mediated iron transport: Crystal structure of FhuA with bound lipopolysaccharide, *Science. 282*, 2215-2220.

13. Qureshi, N., Takayama, K., Mascagni, P., Honovich, J., Wong, R. & Cotter, R. J. (1988) Complete structural determination of lipopolysaccharide obtained from deep rough mutant of escherichia-coli - purification by high-performance liquid-chromatography and direct analysis by plasma desorption mass-spectrometry, *J Biol Chem. 263*, 11971-11976.

14. Agrawal, A. B., Qureshi, Takayama. (1998) 1H and 13C NMR assignments of a lipopolysaccharide obtained from the deep rough mutant of Escherichia coliD31m4 *Magn. Reson. Chem. 36*, 1-7.

15. Martin, N., Hu, H., Moake, M., Churey, J., Whittal, R., Worobo, R. & Vederas, J.

(2003) Isolation, structural characterization, and properties of mattacin (Polymyxin M), a cyclic peptide antibiotic produced by Paenibacillus kobensis M, *J Biol Chem. 278*, 13124-13132.

16. Live, D. H., Davis, D. G., Agosta, W. C. & Cowburn, D. (1984) Observation of 1000-fold Enhancement of ^{15N} NMR via Proton-Detected Multiple-Quantum Coherences: Studies of Large Peptides, *J. Am. Chem. Soc. 106*, 6104-6105.

17. Keller, R. (2004) *The computer aided resonance assignment*, CANTINA Verlag, Goldau.

18. Goddard, T. D. & Kneller, D. G. SPARKY 3 in, University of California, San Francisco

19. Kay, L. E., Ikura, M., Tschudin, R. & Bax, A. (1990) Three-Dimensional Triple-Resonance NMR Spectroscopy of Isotopically Enriched Proteins, *J. Magn. Reson. 89*, 496.

20. Palmer, A. G., Cavanagh, J., Wright, P. E. & Rance, M. (1991) Sensitivity improvement in proton-detected two-dimensional heteronuclear correlation NMR spectroscopy, *J. Magn. Reson. 93*, 151-170.

21. Wittekind, M. & Mueller, L. (1993) HNCACB, a High-Sensitivity 3D NMR Experiment to Correlate Amide-Proton and Nitrogen Resonances with the Alpha-Carbon and Beta-Carbon Resonances in Proteins, *J. Magn. Reson. Ser. B. 101*, 201-205.

22. Olejniczak, E. T., Xu, R. X. & Fesik, S. W. (1992) A 4D HCCH-TOCSY experiment for assigning the side chain 1H and 13C resonances of proteins, *J Biomol NMR. 2*, 655-9.

23. Bax, A., Clore, G. M., Driscoll, P. C., Gronenborn, A. M., Ikura, M. & Kay, L. E. (1990) Practical Aspects of Proton-Carbon-Carbon-Proton Three-Dimensional Correlation Spectroscopy of 13C labelled Proteins, *J. Magn. Reson. 87*, 620-627.

24. Van, d. S., D, Lindahl, E., Hess, B., Groenhof, G., Mark, A. & Berendsen, H. (2005) GROMACS: Fast, flexible, and free, *J Comput Chem. 26*, 1701-1718.

25. Hassinen, T. & Perakyla, M. (2001) New energy terms for reduced protein models implemented in an off-lattice force field, *J Comput Chem. 22*, 1229-1242.

26. Schüttelkopf, A. W. & van Aalten, D. M. F. (2004) PRODRG - a tool for high-throughput crystallography of protein-ligand complexes, *Acta Crystallogr. D Biol. Crystallogr. 60*, 1355-1363.

27. Oostenbrink, C., Villa, A., Mark, A. E. & Gunsteren, W. F. V. (2004) A biomolecular force field based on the free enthalpy of hydration and solvation: The GROMOS force-field parameter sets 53A5 and 53A6, *Journal of Computational Chemistry. 25*, 1656-1676.

28. No, K. T., Grant, J. A. & Scheraga, H. A. (1990) Determination of net atomic charges using a modified partial equalization of orbital electronegativity method. 1. Application to neutral molecules as models for polypeptides, *The Journal of Physical Chemistry. 94*, 4732-4739.

29. Gasteiger, J. & Marsili, M. (1980) Iterative partial equalization of orbital electronegativity--a rapid access to atomic charges, *Tetrahedron. 36*, 3219-3228.

30. Galanos, C., Luderitz, O. & Westphal, O. (1969) A new method for extraction of R-Lipopolysaccharides *Eur. J. Biochem. 9*, 245-249.

31. Shuker, S., Hajduk, P., Meadows, R. & Fesik, S. (1996) Discovering high-affinity ligands for proteins: SAR by NMR., *Science. 274*, 1531-4.

32. Clore, G. M. & Gronenborn, A. M. (1982) Theory and Applications of the Transferred Nuclear Overhauser Effect to the Study of the Conformations of Small Ligands Bound to Proteins, *J. Magn. Reson. 48*, 402-417.

33. Gemmecker, G. (2003) Isotope Filter and Editing Techniques in *BioNMR in Drug Research* (Zerbe, O., ed) pp. 373-390, Wiley-VCH, Weinheim.
34. Blommers, M. J. J. & Rüdisser, S. (2003) NMR of Weakly Binding Ligands in *BioNMR in Drug Research* (Zerbe, O., ed) pp. 355-372, Wiley-VCH, Weinheim.
35. Marchioro, C., Davalli, S., Provera, S., Heller, M., Ross, A. & Senn, H. (2003) Experiments in NMR-Based Screening in *BioNMR in Drug Research* (Zerbe, O., ed) pp. 321-340, Wiley-VCH, Weinheim.
36. Vuister, G. W. & Bax, A. (1992) Resolution enhancement and spectral editing of uniformly C-13-enriched proteins by homonuclear broad-band C-13 decoupling, *J. Magn. Reson. 98*, 428-435.
37. Grzesiek, S. & Bax, A. (1992) Improved 3D triple-resonance NMR techniques applied to a 31-kDa protein *J. Magn. Reson. 96*, 432-440.
38. Jarvet, J., Zdunek, J., Damberg, P. & Gräslund, A. (1997) Three-dimensional structure and position of porcine motilin in sodium dodecyl sulfate micelles determined by ^{1H}NMR, *Biochemistry. 36*, 8153-63.
39. Papavoine, C. H., Konings, R. N., Hilbers, C. W. & van de Veen, F. J. (1994) Location of M13 Coat protein in sodium dodecyl micelles as determined by NMR, *Biochemistry. 33*, 12990-12997.
40. Koch, P., Frank, J., Schuler, J., Kahle, C. & Bradaczek, H. (1999) Thermodynamics and structural studies of the interaction of Polymyxin B with deep rough mutant lipopolysaccharides, *J Colloid Interf Sci. 213*, 557-564.
41. Bhunia, A., Domadia, P. N. & Bhattacharjya, S. (2007) Structural and thermodynamic analyses of the interaction between melittin and lipopolysaccharide, *Biochimica et biophysica acta*, .
42. Gruber, A., Mancek, M., Wagner, H., Kirschning, C. & Jerala, R. (2004) Structural model of MD-2 and functional role of its basic amino acid clusters involved in cellular lipopolysaccharide recognition, *J Biol Chem. 279*, 28475-28482.
43. Ferguson, A., Welte, W., Hofmann, E., Lindner, B., Holst, O., Coulton, J. & Diederichs, K. (2000) A conserved structural motif for lipopolysaccharide recognition by procaryotic and eucaryotic proteins, *Structure. 8*, 585-592.
44. Wimley, W. C. & White, S. H. (1996) Experimentally determined hydrophobicity scale for proteins at membrane interfaces, *Nature Struct. Biol. 3*, 842-848.

Tables

Table 1: Carbon and Proton Chemical Shifts of LPS from the deep-rough mutant of *E. coli*, 1mM LPS in 300mM DPC, pH= 5.8; 40mM MES, T=310K

Sugar moieties				
	GlcN-A	GlcN-B	Kdo-C	Kdo-D
H 1	5.346	4.636		
H 2	3.912	3.854		
H 3	5.136	5.066	1.879	2.079
H 3			1.929	1.707
H 4	3.675	3.872	4.066	4.000
H 5	4.037	3.644	4.067	3.961
H 6	4.022	3.403	3.644	3.547
H 6	3.828	3.722		
H 7			3.866	3.909
H 8			3.848	3.913
H 8			3.591	3.687
C 1	95.56	104.41		
C 2	54.07	55.57		
C 3	75.42	76.07	35.32	36.61
C 4	68.62	74.81	70.57	68.04
C 5	73.36	76.23	66.51	68.39
C 6	70.53	64.76	73.60	74.53
C 7			71.90	71.83

Lipid chains						
	LM		MM			
	L	HM2	M	HM1	HM-4	HM-3
HA	2.306*	2.562	*	2.643	2.312	2.383
HA		2.585	*	2.670	2.366	2.433
HB	1.563*	5.283	*	5.128	3.890	3.910
HC	1.210*	1.519	*	1.548	1.303	1.383
HC		1.559	*	1.601	1.380	1.471
HD					1.281	1.392
CA	36.51*	43.15	*	41.10	45.38	44.18
CB	27.34*	72.37	*	72.50	69.77	69.61
CG	31.50*	37.18	*	36.25	38.30	40.30
CD					27.98	28.08

* Resonances that cannot be distinguished between L and M because of overlap

Table 2: Intermolecular NOEs as detected in ^{13}C-filtered NOESY experiments performed with ^{13}C-labelled LPS and unlabelled PMX-B and –E.

Polymyxin-B						
HM4-H(Cα)	Leu7-Hδ1	w		HM3-H(Cγ)	Phe6-Hε	w
HM4-H(Cβ)	Leu7-Hδ1	s		HM3-H(Cδ)	Phe6-Hδ	s
HM4-H(Cδ)	Leu7-Hδ1	w		HM3-H(Cδ)	Phe6-Hε	s
HM4-H(Cδ)	Leu7-Hδ2	s		HM4-H(Cδ)	Phe6-Hδ	s
HM4-H(Cγ)	Leu7-Hδ2	s		HM4-H(Cδ)	Phe6-Hε	w
HM4-H(Cβ)	Leu7-Hδ2	s		HM1-H(Cβ)	Phe6-Hδ	w
HM4-H(Cα)	Leu7-Hδ2	s		HM4-H(Cβ)	Phe6-Hδ	w
HM4-H(Cβ)	Leu7-Hγ	s		HM3-H(Cγ)	Phe6-Hδ	s
HM4-H(Cβ)	Leu7-Hβ2	s		HM4-H(Cα)	Phe6-Hδ	s
HM4-H(Cβ)	Leu7-Hβ1	w		HM4-H(Cγ)	Phe6-Hδ	s
HM3-H(Cβ)	Phe6-Hβ2	w		HM3-H(Cα)	Phe6-Hδ	s
HM1-H(Cβ)	Phe6-Hβ2	w		HM3-H(Cβ)	Phe6-Hδ	s
HM2-H(Cα)	Phe6-Hβ1	w		HM2-H(Cα)	Phe6-Hδ	s
HM4-H(Cα)	DABA8-Hγ	s		HM2-H(Cβ)	Phe6-Hε	w
HM4-H(Cβ)	DABA8-Hγ	w		HM4-H(Cα)	DABA8-HN	w
HM4-H(Cγ)	Phe6-Hε	w				
Polymyxin-E						
HM4-H(Cα)	Leu6-Hδ1	m		HM2-H(Cα)	Leu6-Hδ1	s
HM4-H(Cα)	Leu7-Hβ	m		HM2-H(Cα)	Leu6-HN	m
HM4-H(Cβ)	Leu6-Hγ	m		HM2-H(Cα)	Leu6-Hγ	m
HM4-H(Cβ)	Leu7-Hβ	m		HM2-H(Cβ)	Leu6-Hγ	m
HM4-H(Cβ)	Leu7-Hδ1	m		HM1-H(Cγ)	Dab4-HN	w
HM4-H(Cβ)	Leu7-Hγ	w		HM1-H(Cγ)	Leu6-HN	m
HM4-H(Cβ)	Leu7-HN	m		HM3-H(Cα)	Leu6-Hδ1	m
HM4-H(Cδ)	Leu7-Hβ	w		HM3-H(Cβ)	Leu6-Hδ1	m
HM4-H(Cγ)	Leu6-Hδ1	s		HM3-H(Cγ)	Leu6-Hδ1	m

8.4 Supplementary material

Table S1: Residual intensity of C,H cross peaks of LPS after addition of 4mM Methyl-5-doxylstearate (pH=4.4, 300mM DPC, 40mM Acetate, T 310K)

Unit	Atoms	Int/Int(0)	Unit	Atoms	Int/Int(0)
HM3	Cγ-Hγ(2)	0.36	GlcN A	C2-H2	0.79
HM3	Cγ-Hγ(1)	0.37	GlcN A	C1-H1	0.85
HM1	Cγ-Hγ(2)	0.45	GlcN A	C3-H3	0.87
HM1	Cγ-Hγ(1)	0.45	GlcN A	C6-H6	0.87
HM3	Cα-Hα(1)	0.46	Kdo C	C6-H6	0.89
HM3	Cβ-Hβ	0.46	Kdo C	C8-H8(2)	0.89
HM3	Cδ-Hδ	0.47	Kdo D	C7-H7	0.89
HM2	Cβ-Hβ	0.47	GlcN A	C6-H6	0.90
HM1	Cα-Hα(2)	0.47	Kdo D	C8-H8(1)	0.90
HM4	Cγ-Hγ(1)	0.48	Kdo C	C3-H3(1)	0.90
HM4	Cγ-Hγ(2)	0.50	GlcN B	C4-H4	0.90
HM3	Cα-Hα(2)	0.50	Kdo D	C8-H8(2)	0.90
HM2	Cα-Hα(1)	0.53	Kdo C	C8-H8(1)	0.91
HM2	Cα-Hα(2)	0.53	Kdo C	C4-H4	0.92
HM2	Cγ-Hγ(1)	0.54	GlcN B	C6-H6	0.92
HM4	Cδ-Hδ	0.55	Kdo D	C5-H5	0.92
HM1	Cβ-Hβ	0.55	Kdo C	C5-H5	0.92
HM2	Cγ-Hγ(2)	0.56	GlcN B	C5-H5	0.93
HM4	Cα-Hα(1)	0.56	GlcN A	C5-H5	0.94
HM2	Cβ-Hβ	0.56	Kdo C	C7-H7	0.96
HM4	Cα-Hα(2)	0.57	Kdo C	C3-H3(2)	0.96
GlcN B	C2-H2	0.63	Kdo D	C6-H6	0.96
GlcN A	C4-H4	0.78	Kdo D	C3-H3(1)	0.97
GlcN B	C3-H3	0.78	Kdo D	C4-H4	1.00
GlcN B	C1-H1	0.79			

Table S2: Chemical shift differences (in ppm) of LPS upon adding PMX-B, PMX-E and PMX-M as derived from the chemical shift mapping experiments. Only changes of chemical shifts, which can be accurately measured, are included.

Unit	Atoms	Polymyxin B		Polymyxin E		Polymyxin M	
		Δ ^{13}C	Δ ^{1}H	Δ ^{13}C	Δ ^{1}H	Δ ^{13}C	Δ ^{1}H
GlcN A	C1-H1	0.084	0.015	0.011	0.018	0.017	0.025
GlcN A	C2-H2	0.015	0.036	0.052	0.021	0.024	0.022
GlcN A	C4-H4	0.312	0.039	0.320	0.090	0.383	0.092
GlcN A	C5-H5	0.154	0.002	0.285	0.024	0.341	0.040
GlcN A	C6-H6	0.109	0.081	0.118	0.084	0.233	0.012
GlcN A	C6-H6	0.038	0.024	0.085	0.004	*	*
GlcN B	C1-H1	0.517	0.031	0.714	0.074	0.329	0.051
GlcN B	C4-H4	0.149	0.129	0.156	0.121	*	*
GlcN B	C5-H5	0.168	0.042	0.186	0.040	*	*
Kdo C	C3-H31	0.552	0.002	0.649	0.009	*	*
Kdo C	C3-H32	0.522	0.037	0.603	0.048	*	*
Kdo C	C4-H4	0.271	0.006	0.287	0.008	0.288	0.006
Kdo C	C5-H5	0.174	0.023	0.182	0.027	0.118	0.034
Kdo C	C6-H6	0.269	0.027	0.280	0.026	0.241	0.076
Kdo D	C3-H31	0.199	0.002	0.179	0.001	0.206	0.002

Kdo D	C3-H32	0.174	0.012	0.182	0.006	0.218	0.014
Kdo D	C4-H4	0.051	0.007	0.042	0.009	0.064	0.006
Kdo D	C5-H5	0.037	0.006	0.023	0.005	0.002	0.001
Kdo D	C6-H6	0.022	0.018	0.021	0.020	0.081	0.018
HM4	Cα-Hα	0.061	0.035	0.196	0.054	0.104	0.000
HM4	Cβ-Hβ	0.078	0.031	0.083	0.002	*	*
HM4	Cγ-Hγ	0.013	0.014	0.139	0.000	0.031	0.043
HM3	Cα-Hα	0.000	0.016	0.042	0.022	0.115	0.019
HM3	Cβ-Hβ	0.078	0.029	0.067	0.025	*	*
HM3	Cγ-Hγ	0.048	0.007	0.041	0.025	0.021	0.017
HM2	Cα-Hα	0.371	0.007	0.472	0.015	0.468	0.006
HM2	Cβ-Hβ	0.600	0.120	0.928	0.079	1.068	0.067
HM2	Cγ-Hγ	0.156	0.005	0.080	0.051	0.285	0.027
HM1	Cα-Hα	0.068	0.029	0.001	0.047	0.105	0.052
HM1	Cγ-Hγ	0.026	0.001	0.264	0.006	0.353	0.006

Fig. S3: Selected spectral regions of the constant-time [^{13}C,^{1}H]-HSQC of 0.35mM LPS in 300 mM DPC, 40mM acetate pH=4.4, T=310K. Therein, original peak positions are shown in yellow, marked with cross and labeled for the carbon atom. The LPS signals that are shifted upon addition of equimolar amounts of PMX-B, -E and -M are shown in blue, magenta and green, respectively.

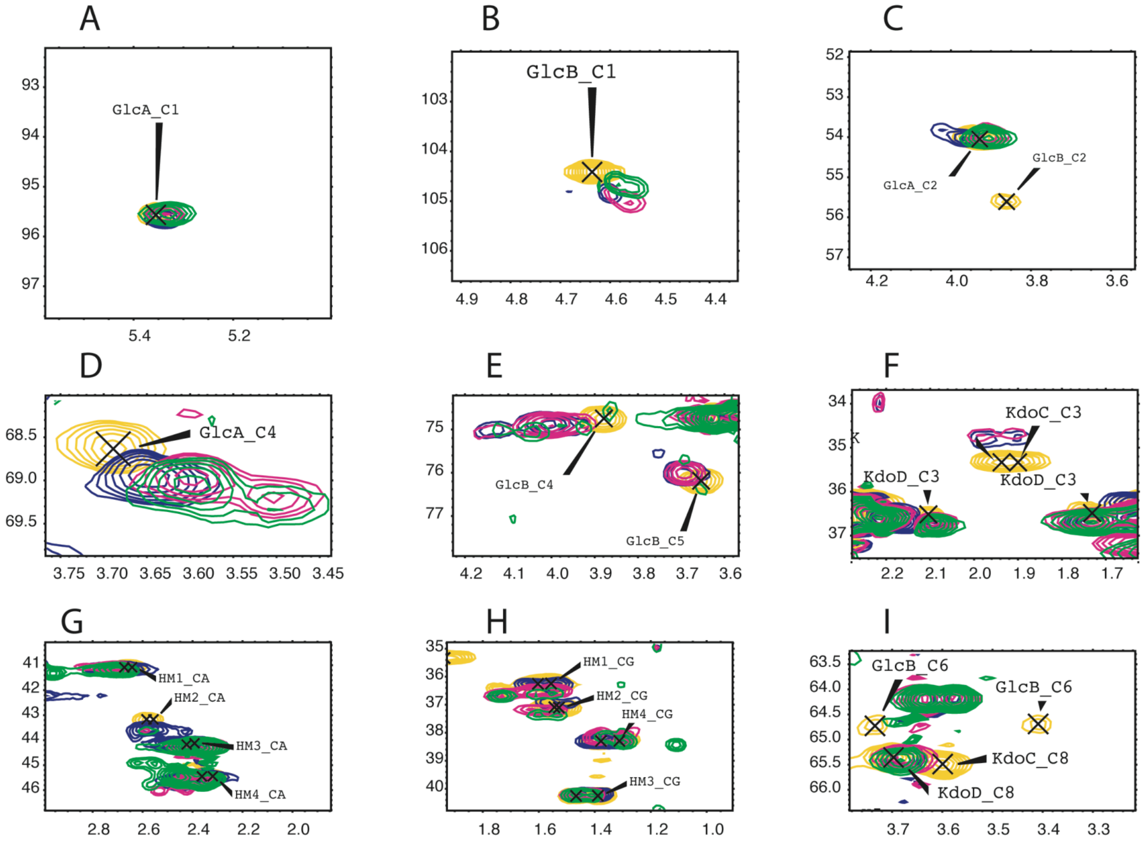


Fig. S4: Elution profile for the HPLC purification of LPS in the ternary solvent mixture showing MALDI MS spectra of selected fractions. The dominant compound of fraction 3 is the form of LPS, which has one phosphate group substituted by pyrophosphate, theoretical MW: 2429.665 for 100% ^{13}C ^{15}N. Fractions 5, 6 and 7 contain the desired form of LPS, theoretical MW: 2349.699.

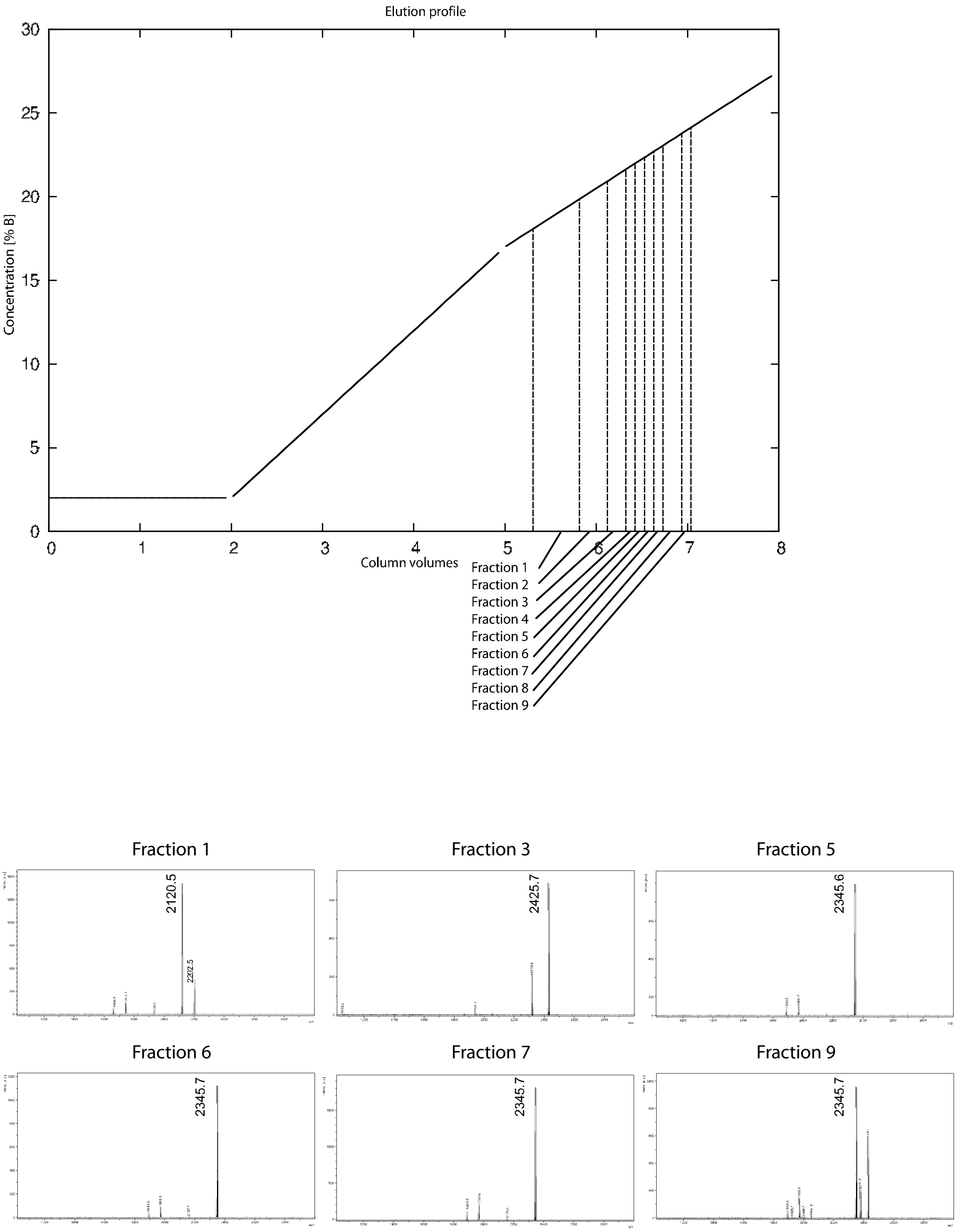


Fig. S5: Anomeric region of the constant-time [^{13}C,^{1}H]-HSQC of LPS. The spectrum is shown close to the noise level to document the purity of the obtained LPS.

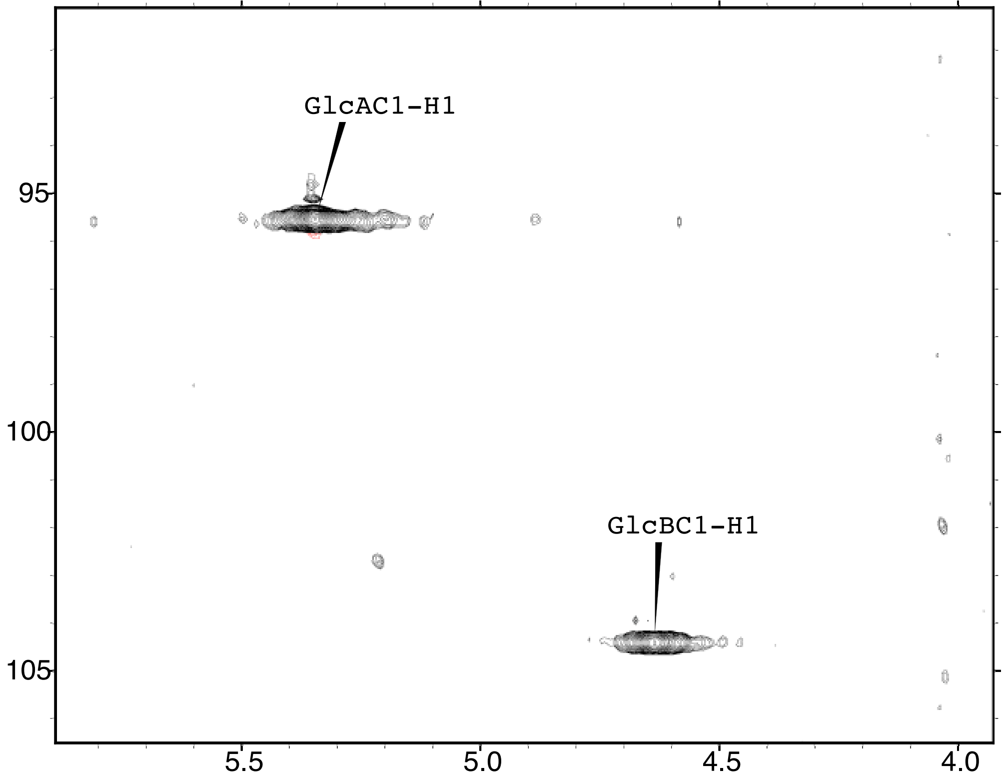


Fig. S6: Differences in LPS chemical shift *changes* when adding PMX between PMX-B vs. PMX-E (left) and PMX-B vs. PMX-M (right).

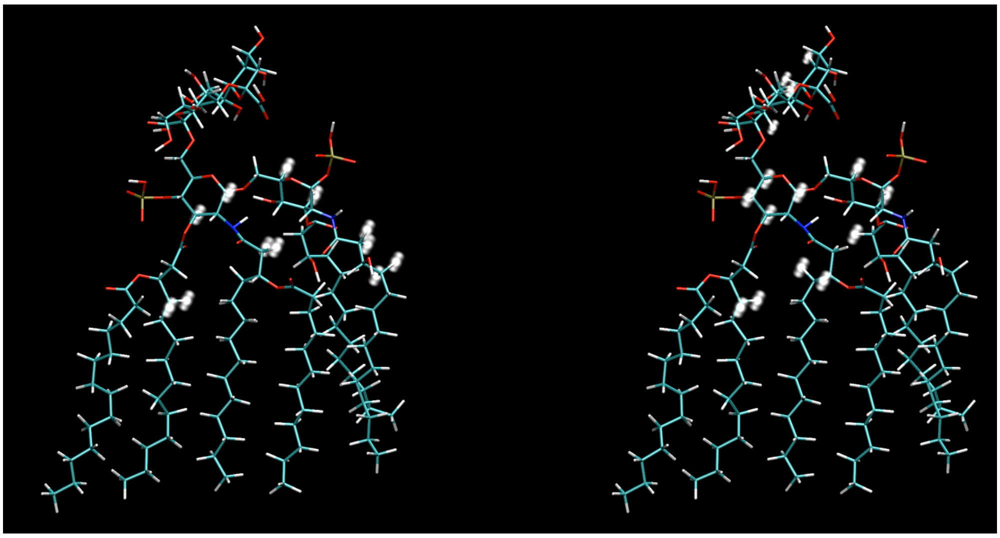


Fig. S7: Structure of PMX-B in the complex with LPS as derived from the MD calculation. Color-coding is according to Figure 6 from the manuscript. Different orientations are shown.

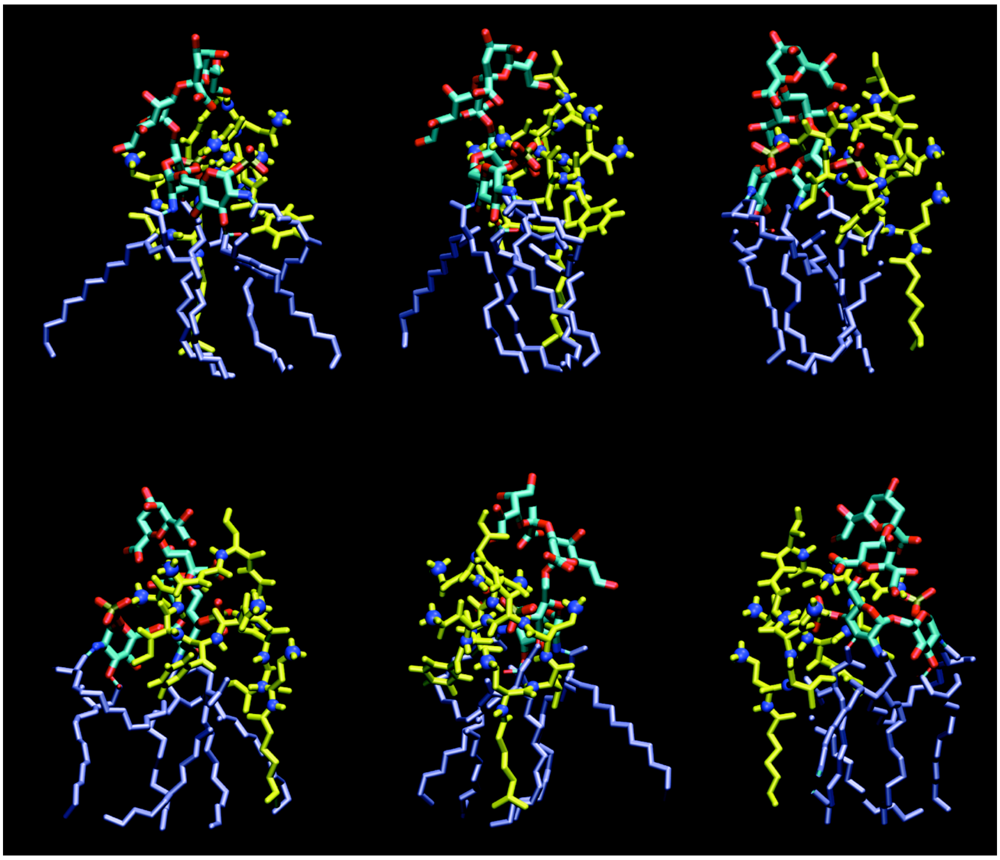


Table S8: Chemical shifts of polymyxin E, 310K, 2mM PMX-E, 300 mM DPC, pH=4.4, 40mM acetate buffer. All chemical shifts were referenced relative to DSS at 310K (FA: fatty acid chain)

	H^N	$H^{'\alpha}$	H^β	Others
FA		2.30	1.56	γCH_2 1.29, δCH_2 1.18; εCH_2 1.10;CH_3 0.85
Dab 1	8.49	4.52	2.07, 2.20	γCH_2 3.05, 3.05; δNH_3^+ -
Thr 2	8.23	4.36	4.30-	γCH_3 1.17; γOH -
Dab 3	8.81	4.36	2.14, 2.24	γCH_2 3.08, 3.08; δNH_3^+
Dab 4	8.38	4.15	1.93, 1.99	γCH_2 3.08, 3.45; δNH 7.62
Dab 5	8.24	4.47	2.04, 2.16	γCH_2 3.00, 3.00; δNH_3^+
Leu 6	8.60	4.25	1.52, 1.63	γCH 1.67; Hδ 0.93, 0.91
Leu 7	8.78	4.31	1.60, 1.67	γCH 1.60; Hδ 0.85, 0.93
Dab 8	7.93	4.40	2.14, 2.28	γCH_2 3.12, 3.12; δNH_3^+ -
Dab 9	8.23	4.67	2.21, 2.28	γCH_2 3.05, 3.05; δNH_3^+ -
Thr 10	7.81	4.11	4.28	γCH_3 1.19; γOH -

Table S9: Chemical shifts of polymyxin B, 310K, 2mM of PMX-B, 300 mM DPC, pH 4.4, 40 mM acetate buffer

	H^N	$H^{'\alpha}$	H^β	others
FA	-	2.30	1.56	γCH_2 1.30; δCH_2 1.19; CH_3 0.91
Dab 1	8.51	4.53	2.09, 2.20	γCH_2 3.05; δNH_3^+ -
Thr 2	8.24	4.35	4.30	γCH_3 1.19; γOH -
Dab 3	8.81	4.38	2.14, 2.24	γCH_2 3.09, 3.09; δNH_3^+ -
Dab 4	8.38	4.13	1.93, 1.93	γCH_2 3.06, 3.45; δNH 7.62
Dab 5	8.20	4.44	1.96, 2.09	γCH_2 2.90, 2.90; δNH_3^+ -
Phe 6	8.68	4.55	2.98, 3.08	δH 7.26, 7.26; εH 7.16, 7.16; ζH -
Leu 7	8.61	4.19	1.35, 1.49	γH 0.88; δCH_3 0.68, 0.75
Dab 8	7.96	4.39	2.14, 2.25	γCH_2 3.10, 3.10; δNH_3^+ -
Dab 9	8.14	4.39	1.91, 2.01	γCH_2 2.85, 2.85; δNH_3^+ -
Thr 10	7.83	4.11	4.26	γCH_3 1.18; γOH -

Table S10: Chemical shifts of polymyxin M, 310K, 2mM PMX-M, 300 mM DPC, pH=4.4, 40mM acetate buffer

	H^N	$H^{'\alpha}$	H^β	others
FA		2.33	1.59	γCH_2 1.33; δCH_2 1.24; εCH_2 1.34; ζCH_2 1.25; QD 0.87; QE 0.87
Dab 1	8.45	4.56	2.22, 2.12	γCH_2 3.10, 3.10; δ NH_3^+ -
Thr 2	8.37	4.25	4.23	γCH_3 1.25; γOH -
Dab 3	8.80	4.45	2.27, 1.71	γCH_2 3.12, 3.12; δNH_3^+ -
Dab 4	8.15	4.27	2.00, 2.00	γCH_2 3.42, 3.16; δNH 7.8
Dab 5	8.34	4.59	2.23, 2.09	γCH_2 3.08, 3.08; δNH_3^+ -
Leu 6	8.77	4.45	1.69, 1.64	γH 1.69; δCH_3 0.99, 0.93
Thr 7	8.56	4.37	4.44	γCH_3 1.19; γOH -
Dab 8	8.35	4.36	2.31, 2.27	γCH_2 3.19, 3.19; δNH_3^+ -
Dab 9	8.79	4.29	2.29, 2.17	γCH_2 3.11, 3.11; δNH_3^+ -
Thr 10	7.94	4.15	4.26	γCH_3 1.22; γOH -

Table S11: Relative attenuations of signals from PMX-B, -E or –M after addition of 5 equiv. of Methyl-5-doxylstearate. The experiments used 4mM unlabelled PMX-B, PMX-E and 0.2mM ^{13}C,^{15}N-labelled PMX-M in 300mM DPC in Acetate buffer pH 4.4, 310K.

Residue Number	PMX-B	PMX-E	PMX-M
1	0.11	0.17	0.03
2	0.71	0.77	0.50
3	0.84	0.80	0.73
4	0.90	0.73	0.75
4(Gamma)	0.50	0.77	0.47
5	0.74	0.84	0.56
6	0	0	0.02
7	0	0	0.22
8	0.57	0.92	0.93
10	0.81	0.69	0.68

NMR sample preparation protocol for the shift mapping experiments:

Several milligrams of pure lyophilized LPS powder were accurately weighted, dissolved in deionized water. A small amount of deuterated DPC was added, which allowed to prepare higher concentrations of LPS. This stock solution was aliquoted and lyophilized for further use. (The additional amount of DPC introduced this way was negligible with respect to the amount of DPC in the final NMR sample).

Several milligrams of pure lyophilized polymyxins were weighted and stock solutions were prepared by dissolving the peptide in the acetate buffer, which was used for measurement of NMR spectra.

For NMR measurements, an aliquoted amount of LPS was dissolved in the corresponding DPC/acetate solution. Good homogeneity of the sample was achieved by brief (10 s) sonication. In the interaction studies, the amount of acetate buffer, in which DPC was dissolved, was reduced by the volume of added polymyxin solution.

In separate experiments we checked that LPS and polymyxins in concentrations used for the sample preparation did not change the pH of the sample. The pH of the sample was not adjusted after the mixing to avoid possible contaminations from the electrode electrolyte components.

Protocol and details for the MD simulation:

Setup

In the first step, a set of structures was generated using a simulated annealing protocol. All of the experimentally observed NOEs were incorporated as distance restraints in this calculation. Simulated annealing *in vacuo* allowed fast generation of a diverse set of starting structures that fulfill the experimental constraints. The electrostatic interactions were scaled by a dielectric constant of 20, which roughly mimicks the interface of water/membrane environment.

The set of annealed structures was clustered, so that the most abundant conformations can be selected. A representative conformer from each of the resulting 10 clusters was transferred to the simulation box that included a DPC micelle for further simulation. LPS-polymyxin complexes were placed into vicinity of the equilibrated DPC micelle consisting of 54 DPC molecules. The subsequent solvation used the same number of water molecules in each case. An artificial atom was placed in the center of the simulation box, which was position-restrained by a strong force. Distance restraints were between this atom and the atoms at the center of the aliphatic chain of each DPC molecule as well of the LPS lipid chains were applied. The restrained distance corresponds to the approximate average distance of these atoms from the center of the micelle. The distance restraints were time averaged giving the system sufficient motional freedom. The experimental distance restraints between LPS and polymyxin were strongly kept to preserve the structure of LPS-polymyxin complex. The system was driven close to equilibrium within half a nanosecond MD simulation. Each system was then placed into a smaller simulation box and solvated by an accordingly smaller number of water molecules.

The distance restraints were then released and the system was allowed to equilibrate within a 2 ns simulation, after which a short production MD simulation followed.

Improvement

To further improve agreement with experimental data a final structure from the trajectory described above that gave the best score was used to create new set of diverse structures. Therein, simulated annealing cycles were calculated, in which the DPC molecules were very weakly position restrained while atoms of LPS and polymyxin were allowed to move freely except for the terminal atoms of aliphatic chains of LPS. The water was removed and replaced by a implicit solvent model. This setup generated a set of diverse conformations, in which steric interactions of atoms of LPS or PMX-B with DPC as well as electrostatic interactions were reasonably taken into account. At the end of this procedure a set of systems was established that included the DPC micelle into which the LPS-polymyxin complex was embedded.

During this step, distance restrains were exclusively applied to specific interactions involving those aliphatic LPS atoms for which significant changes were observed by chemical shift mapping. Since the interaction partner atom of PMX-B was unknown, various possibilities were tested. In all cases, long-range distance restraints of Phe6 ring atoms to atoms of the fatty acids were used. From a set of 450 annealed structures 45 were selected according to the occurrence of experimentally verified interactions. These were then further subjected to a short (120 ps) equilibration simulation with explicit solvent. Five trajectories were finally computed without restraints and a representative snapshot from the most stable trajectory is shown in Figure 6 in the main text.

Analysis

Each of the resulting MD trajectories was assessed in several steps. First, RMS deviations with respect to the start of the production trajectory were calculated to monitor, whether the MD trajectory is stable. All atoms of LPS and polymyxin were used for RMSD calculation. These trajectories were derived from the conformations based on NOE contacts, therefore another experimental information was required to evaluate, how they reflect the experimental system. Average distances between each atom of LPS with each atom of the polymyxin were calculated over the MD trajectory. Those distances were selected for which participating atoms displayed changes in the chemical shift mapping experiments.

All the analysis was done using tools from the GROMACS MD package and by own scripts. Structures were visually inspected by VMD, which was also used to prepare figures.

Simulation parameters

Vacuo simulated annealing

All the interactions were computed without any cutoff for reasons of computational stability, the dielectric constant was set to 20, the integration step to 0.5 fs, the force constant for distance restraints to 1000 kJ mol-1 nm-2. The annealing cycle had the following profile: 80 ps at 1200K, linear cooling to 298 over 5 ps, 55 ps at 298 K, linear cooling to 0 K over 50 ps and the cycle was closed by linear heating to 1200K over a time of 10ps.

Simulated annealing with implicit solvent

The atoms of DPC molecules were position restrained with a force constant of 20, terminal atoms of LPS aliphatic chains with force constant 100. The dielectric constant was set to 80 and the GBSA implicit solvent model as implemented in GROMACS 4.0 was used. The simulated annealing cycle was shortened 4 times to save computer time.

Equilibration and production MD

For initial equilibration the system of DPC, LPS and polymyxin was solvated with 12574 SPC water molecules. Equilibration after simulated annealing with implicit solvent and the production MD was performed with 7295 water molecules. One Cl^- ion was added to neutralize the net charge. In the initial equilibration and production MD, the integration step was set to 1 and 2 fs, respectively. The simulations were performed in the NPT ensemble using periodic boundary conditions. Long-range electrostatics were treated by using Particle-mesh Ewald summation. The temperature was weakly coupled (coupling time 0.1 ps) to T=310K. Bond lengths were constrained using the LINCS algorithm. The pressure was weakly coupled (coupling time 0.5 ps, compressibility 4.5e-5 bar) to 1 bar. The force constant of distance restraints during the equilibration was set to 5000.

Curriculum Vitae

Personal details

Surname: MAREŠ
First name: Jiří
Date of birth: 7.5.1980
Nationality: Czech

Education

1994–1998	High school: Gymnasium Kladno, Czech republic
1999–2004	Institute of Chemical Technology Prague, specialization biochemistry
2003–2004	**Diploma Thesis** Institute of Chemical Technology Prague. Supervision of Dr. Ing. Richard Hrabal, Prof. Tomáš Ruml subject: *Study of Structure and Interactions of Structural Proteins of Retroviral Capside* degree: Engineer of chemistry
2004/12–2008/12	PhD. position at the Department of Organic Chemistry at University of Zürich (Prof. Dr. Oliver Zerbe) subject: *Structure, Folding and Interactions of Membrane-associated Biomolecules studied by NMR*

PUBLICATIONS

1. **Probing the formation of stable tertiary structure in a model miniprotein at atomic resolution: Determinants of stability of a helical hairpin**
   Alexey Neumoin, Jiří Mareš, Mirjam Lerch-Bader, Reto Bader, Oliver Zerbe; *Journal of The American Chemical Society 129 (28)*: 8811-8817, **2007**

2. **A model for cell-surface-exposed carbohydrate moieties suitable for structural studies by NMR spectroscopy**
   Jiří Mareš, Jan Müller, Audrone Skirgailiene, Alexey Neumoin, Carole A Bewley, Richard R Schmidt, Oliver Zerbe; *ChemBioChem 7 (11)*: 1764-1773, **2006**

3. **Interactions of lipopolysaccharide and polymyxin studied by NMR spectroscopy**
   Jiří Mareš, Sowmini Kumaran, Marina Gobbo, Oliver Zerbe; in revision *Journal of Biological Chemisty*

4. **Structural and biochemical studies on procaspase-8: new insights on initiator caspase activation**
   Nadine Keller, Jiří Mareš, Oliver Zerbe, Markus Grütter; in press *Structure*

5. **Studies of unfolding of PYY in water and methanol by NMR**
   Jiří Mareš, Reto Walser, Alexey Neumoin, Chao Zou, Oliver Zerbe; manuscript in preparation.

Acknowledgements

First of all, I would like to express my thanks to my supervisor, Prof. Oliver Zerbe for the friendly supervision during the work on my thesis, for many skills I learned from him during the work on a variety of projects and for giving me an opportunity to work often on those part of shared projects which best fitted my scientific interests.

I would also like to thank to following people who directly or indirectly helped me to complete this work:
-Prof. Wilfred van Gunsteren for co-refereeing my thesis
-Prof. John Robinson and Prof. Kim Baldridge for support and kind provision of laboratory and computational resources

-Simon Jurt and Nadja Bross from NMR service by far not only for their great work

-MS- service, with Dr. Laurent Bigler in head, for excellent service they provide

-all members of Prof. Zerbe's group. Sowmini Kumaran and Chao Zou for a lot of human support during the difficult beginning of my thesis in Zurich, Sowmini also for the ability to share with me a project, Stefan Markovic and, on longer term, Reto Walser for silent support when things went fine and active support whenever some help was needed, Reto also for correcting the German translation of the thesis summary,
-remaining friends from the group: Nadine and Alexey

-Svava Kim Wetzel and Christina Ewald, who, during the final stages of my thesis, largely contributed to the fact, that leaving this scientific group is also leaving the great friends, special thanks belong to Christina, for tens of hours spent on thorough reading of the manuscript of my thesis, coming with bright ideas, helpful discussions and suggestions,

-Krystyna Patora, who, in addition to scientific discussions, contributed to pleasant time in Zurich from the very first day and also introduced me to other people, who became my close friends, especially Renata Windak and Dariusz Ekononomiuk, to whom I also express my thanks, Darek also for scientific discussions
- Nelli Rahanyan, who shared a lot of understanding with me,
- Radovan Fišer for many scientific discussions and computational resources

- Prof. Robinson's group:
Katarina, Nina, Jessica, Jasmin, Tina, Roba, Frédéric, Jimi, Arin, Emad, Annelies, Patrick, Tanja, Markus, Dr. Katja Zerbe and Dr. Kerstin Moehle, whom I have been meeting everyday and who contributed to smiley start of the working days, Dr. Moehle also for many scientific discussions,

Marie-Therese Bohley, Dr. Christa Werner, Cornelia Moor, Salomé Fässler, Dr. Roland Kunz for the combination of perfect administration with kind and friendly approach to people,

-members of my broader Swiss family, naming at least - Ania, Ala, Marek, Matěj, Jakub, Sini, Lukasz, Marta, Philippe, Ruta, Miriam, Simone, Ruth, Dominika, Iza, Jana, Masoud, Ilona - for being Here.

Last, but biggest thanks to my parents and sister with family for the greatest possible support over the distance of 800km